SPEECH CODING BY COMPLEX AM AND FM SIGNAL MODELS

By

Deshraj Singh

DEPARTMENT OF ELECTRICAL ENGINEERING

Indian Institute of Technology Kanpur

FEBRUARY, 2002

SPEECH CODING BY COMPLEX AM AND FM SIGNAL MODELS

A Thesis Submitted
in Partial Fulfillment of the Requirement for the Degree of

Master of Technology

By
DESHRAJ SINGH

To the
DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY KANPUR
FEBRUARY 2002

CERTIFICATE

It is certified that the work contained in the thesis entitled "SPEECH CODING BY COMPLEX AM AND FM SIGNAL MODELS" by Deshraj Singh has been carried out under my supervision and that this work has not been submitted elsewhere for degree.

Pradip Sircar

Dr. Pradip Sircar
Assistant Professor
Department of Electrical Engineering
Indian Institute of Tecnology
Kanpur

February, 2002

Abstract

Complex AM and FM signal models can be used for representation of non –stationary signals such as speech [1,2,3]. Complex AM signal model has been found to be suitable for sustained voiced speech phonemes [1,3],while Complex FM signal model can be used for representation of sustained unvoiced speech phonemes [2,3]. But this type of classification is not appropriate, in this study it is shown that the phonemes having most of their energy in low frequency region can be fitted by complex AM model ,while for those having most of their energy in high frequency region Complex FM model is suitable. Also in sustained vowel and consonant sounds the gain of the signal is constant so Complex AM and FM models can directly give the parameters. but in natural spoken speech signal the gain varies with time .This study considers the time-varying nature of speech signal gain and explains principle of parameter estimation by these two models by making gain of speech signal constant .Time varying gain of the speech is estimated and fitted by polynomial model .The parameters of constant gain speech signal and coefficients of polynomial are coded.

Acknowledgement

I am deeply indebted to **Dr. Pradip Sircar**, for his inspiring guidance and encouragement at every *stage of this thesis . He provided me with numerous* constructive suggestions, without which this work would not have been possible. Above all , I am grateful to him for the patience displayed during those numerous discussions.

I am thankful to Dr. S. Umesh, for his help . He explained me various concepts related to speech coding , which made my work very easy. I am also thankful to S.V. Bharat Kumar, Rajesh Sharma and Harsha for their great help during this period. I pay special thank to Rohit Sinha, for providing me speech data from Speech Lab. *for simulation studies.*

During my stay at IIT, all my friends Subrata Maiti, Sachin Srivastava ,Sachin Nagar,Pankaj Kumar, Ketan ,Sampurnand Mishra, Shashi Kant, Rajesh Shrama,Anil kumar have given me all support I needed to develop my self as good human being. I am grateful to my teachers which not only taught me courses but also a way to be a better person. .I am very grateful to Dr. R.K.Bansal who taught me the way of hard work and thinking. This acknowledgement will remain incomplete without acknowledging the continued support from Pranay Ji. He is continued source of inspiration for me.

And at last but not the least, I salute the one who is manifesting himself in whole of the universe. That one has made me able to fulfill this job.

List of Figures

List of Tables

Chapter 1

Introduction

One approach to the problem of representation of speech signal is to use the speech production model in which speech is viewed as the result of passing a glottal excitation waveform through a time-varying linear filter that models the resonant characteristics of the vocal tract. It is assumed that glottal excitation can be in one of two possible states quasi-periodic pulses (during voiced speech), or random noise (during unvoiced speech). In this model the basic speech parameters, e.g., pitch, formants, spectra, vocal tract area functions etc. are estimated and coded for speech compression.

In above model the notion of quasi-stationarity is applied, and the identification of model is performed over short data segments [12,13]. In this case, a compromise is needed between the faithfulness of the model in representing the details of signal and accuracy of estimation of the parameters of the model.

Low rate speech coding by sinusoidal model has been addressed [12] in this model the glottal excitation is modeled by sinusoidal components arbitrary amplitudes, frequencies and phases. This model is also frame based, each frame is of 10 to 20 ms. And generally the length of one phoneme is above 40 ms. Which means that for a single phoneme one has to

take 4 or 5 frames, in other words 4 or 5 parameter sets are to be taken for one phoneme in general.

Multicomponent AM and FM signal models are developed for representation of voiced and unvoiced speech phonemes respectively [1,2,3].

To estimate the parameters of the multicomponent complex AM signal model, the accumulated autocorrelation function (AACF) of speech data is computed by taking the sum of time dependent autocorreation functions (ACFs) over an assigned time frame. The PSD plot of AACF is then used to obtain carrier and modulating frequencies. Once the frequencies are known the amplitude and modulation index parameters of the fitted model can be found by solving linear estimation problem. [1]

The above technique of frequency estimation was not suitable for fast on-line processing of data .So a new technique, which involves fitting the AACF sequence in a linear prediction model, was demonstrated. The Zeros of the prediction error filter (PEF) are used to estimate the carrier and modulating frequency [3]. This enabled the fast on-line automatic processing of speech data.

For estimation of frequency parameters of the multicomponent complex FM signal model, a sequence of product functions of the signal samples is computed and then processed to obtain the autoregressive PSD of underlying process. This PSD plot is studied in conjunction with DFT plot of the data to obtain carrier and modulating frequencies. Modulation indices are obtained from magnitude plot of DFT.the remaining parameters can be estimated by linear estimation [2]

The above technique was again not suitable for on-line processing of data Moreover, determination of modulation indices from magnitude plot of DFT involves separation of plot into its individual components, which makes the parameter estimation problem more difficult. So a new technique which estimates the frequencies by linear production model was demonstrated [3] But in all these methods explained in [1,2,3] of parameter estimation and *model fitting of speech data the time-varying gain or variance of speech is* not considered and model is fitted for sustained vowel and consonant sounds.

In this study Time-varying gain of speech is estimated and the original speech signal is divided by this estimated gain to make resulting signal of constant gain. Then complex AM or FM models are applied on this constant gain speech, to estimate parameters.

Gain function is fitted by polynomial model and coefficients of polynomial are estimated in least square sense. The speech parameters and coefficients of gain are coded to check the potential of speech compression by models proposed in [1,2,3].

Chapter 2

Complex AM Signal Model

2.1 The Model

In this section the complex AM signal model as proposed and explained in [1] is introduced. The discrete-time complex random process x[n] consisting of M signal-tone amplitude modulated signal is represented by

$$x(n) = \sum_{i=1}^{M} A_i \left[1 + \mu_i\, e^{jv_i nT}\right] e^{jw_i nT}\, e^{j\phi_i} \tag{2.1}$$

Where A_i is the carrier amplitude of constituent signal, μ_i is the modulation index, ω_i is the carrier angular frequency, v_i is the is the modulating angular frequency, ϕ_i is the independent and identically distributed (i.i.d) random phase, and T is the sampling interval. It is assumed that the random variable ϕ_i is uniformly distributed over $[0,2\pi]$.

The complex sequence $x(n)$ may be utilized to express the time-dependent autocorrelation function (ACF)

$$r_x(n,k) = E\{x^*(n)x(n+k)\} \tag{2.2}$$

where E stands for the expectation operator, and * denotes complex conjugation.

Substitution for $x^*(n)$ and $x(n+k)$, and subsequent evaluation of the right hand side of Eqn. (2.2) yields

$$r_x(n,k) = \sum_{i=1}^{M} A_i^2 \left[1+\mu_i e^{-j\nu_i nT}\right]\left[1+\mu_i e^{j\nu_i (n+k)T}\right] e^{j\omega_i kT} \qquad (2.3)$$

Where the following result of expectation is utilized: -

$$E\left\{e^{j(\phi_i-\phi_l)}\right\} = \begin{cases} 1 & for \quad \phi_i = \phi_l \\ 0 & for \quad \phi_i \neq \phi_l \end{cases}$$

The accumulated autocorrelation function (AACF) $C_x(k)$ is computed by taking the sum of ACF, $r_x(n,k)$ over a fixed time frame $[n_1, n_2]$,

$$C_x(k) = \sum_{n=n_1}^{n_2} r_x(n,k) \qquad (2.4)$$

Evaluating the right hand-side of Eqn. (2.4) one gets

$$C_x(k) = \sum_{i=1}^{M} B_i e^{j\omega_i kT} + \sum_{i=1}^{M} C_i e^{j\nu_i kT} \qquad (2.5)$$

Where

$$B_i = A_i^2 \left[(n_2 - n_1 + 1) + \mu_i \sum_{n=n_1}^{n_2} e^{j\nu_i nT} \right]$$

and

$$C_i = A_1^2 \mu_i \left[\mu_i (n_2 - n_1 + 1) + \sum_{n=n_1}^{n_2} e^{j\nu_i nT} \right]$$

Eqn. (2.5) rewritten provides the final expression for the time-independent AACF $C_x(k)$ as follows: -

$$C_x(k) = \sum_{i=1}^{2M} R_i e^{j\sigma_i kT} \tag{2.6}$$

where

$$R_i = \begin{cases} B_i & for \quad i = 1,\ldots,M \\ C_{i-M} & for \quad 1 = M+1,\ldots,2M \end{cases}$$

and

$$\sigma_i = \begin{cases} \omega_i & for \quad i=1,\ldots,M \\ \omega_{i-M} + \nu_{i-M} & for \quad i=M+1,\ldots,2M \end{cases}$$

It is explicit from Eqn. (2.6) that the AACF $C_x(k)$ comprises of sinusoids with angular frequencies same as the carrier angular frequencies for the modeled signal $x(n)$.

2.2 Model fitting and Parameter estimation

Starting with the sampled sequence [1] $\{\ y(n); n=0,1,\ldots,N-1\ \}$, which is to be modeled, one first computes the set of AACFs $\{C_y(k);\ k = -J,\ldots,0,\ldots J\}$ given by

$$C_y(k) = \sum_{n=n_1}^{n_2} y^*(n)y(n+k) \qquad (2.7)$$

where the length of the time frame is taken as $n_2 - n_1 + 1 = N - 2J$. Note that the sequence $y(n)$ may be considered as a single observation or sample of the discrete-time random process $x(n)$. Furthermore, since the concept of ergodicity does not apply for non-stationary signals, one has no choice but to drop the expectation operator from Eqn. (2.7).

The sequence $C_y(k)$ is fitted into a linear prediction model of AR (p) process to obtain a prediction error filter (PEF) defined as follows: -

$$A(z) = 1 + a[1]z^{-1} + a[2]z^{-2} + \cdots + a[p]z^{-p} \qquad (2.8)$$

where $a[1], a[2], \cdots, a[p]$ are the linear prediction coefficients (AR) parameters to be determined. Zeros of $A(z)$ will be used to determine the angular frequencies of the model.

In order to estimate the AR parameters, Modified Covariance Method [4] has been employed. This method appears to yield statistically stable spectral estimates with high resolution. For any input $\chi(n)$ it can be written in matrix form as follows: -

$$\begin{bmatrix} C_{xx}[1,1] & C_{xx}[1,2] & \cdots & C_{xx}[1,p] \\ C_{xx}[2,1] & C_{xx}[2,2] & \cdots & C_{xx}[2,p] \\ \vdots & \vdots & & \vdots \\ C_{xx}[p,1] & C_{xx}[p,2] & C_{xx}[p,3] & C_{xx}[p,p] \end{bmatrix} \begin{bmatrix} \hat{a}[1] \\ \hat{a}[2[\\ \vdots \\ \hat{a}[p] \end{bmatrix} = - \begin{bmatrix} C_{xx}[1,0] \\ C_{xx}[2,0] \\ \vdots \\ C_{xx}[p,0] \end{bmatrix} \tag{2.9}$$

where

$$C_{xx}(j,k) = \frac{1}{2(N-p)} \left(\sum_{n=p}^{N-1} \chi^*(n-j)\chi(n-k) + \sum_{n=0}^{N-1-p} \chi(n+j)\chi^*(n+k) \right)$$

where N is the number of data points.

The matrix in Eqn. (2.9) has been decomposed using Singular Value decomposition (SVD) technique [6,8]. Singular values which are comparatively small in magnitude are set to zero before obtaining a solution for linear equations given in Eqn.(2.9) Zeroing of smaller singular values adds inherent noise immunity to estimate technique[7]. Once the AR parameters are obtained, we determine zeros of PEF defined by Eqn. (2.8). From these zeros, the carrier and modulating angular frequencies of the signal $x(n)$ are obtained.

Identification of modulating angular frequencies from the Zeros of PEF is necessary, before finding carrier and modulating angular frequencies. For this the residues R_i are computed for all frequencies of the

AACF sequence. It has been shown [1] that for an unmodulated carrier, the residue is real, whereas the residues are complex for the modulated frequencies. This feature is employed to identify the unmodulated carrier frequencies. Rests of the peaks are considered in pairs taking two adjacent peaks for each modulated carrier. Once all frequencies are known. The amplitude, phase and modulation-index parameters of the fitted model can then be obtained in the next stage by solving a linear estimation problem as presented in the following paragraph.

Considering the inclusion of unmodulating carrier frequencies in the model, the random sequence $x(n)$ can be modeled as

$$x(n) \quad = \quad \sum_{i=1}^{M} A_i\left[1+\mu_i e^{j\nu_i T}\right]e^{j\omega_i nT}e^{j\phi_i} \ + \sum_{i=M+1}^{L} A_i e^{j\omega_i nT}e^{j\phi_i} \qquad (2.10)$$

where the $(L-M)$ unmodulated carrier angular frequencies are include, and the last term of the above equation constitutes the stationary part of the discrete-time process.

The discrete-time signal $y(n)$ is fitted into the complex AM signal model of Eqn. (2.1), which is rewritten as

$$y(n) = \sum_{i=1}^{M} A_{ci}\,\xi_i^n + \sum_{i=1}^{M} A_{ci}\,\mu_i\,\xi_i^n\,\zeta_i^n \qquad (2.11)$$

for $n = 0, \cdots, N-1$,

where $A_{ci} = A_i e^{j\phi_i}$ is the unknown complex amplitude of the carrier, μ_i is the unknown modulation index, and $\xi_i = e^{j\omega_i T}, \zeta_i = e^{j v_i T}$ are the parameters which can be computed from the estimated ω_i *and* v_i values.

For $(L-M)$ unmodulated frequencies $\zeta_i = 1$, hence $\xi_i \zeta_i = \xi_i$, avoiding the ill conditioning, the equation (2.11), can be written as follows:-

$$y(n) = \sum_{i=1}^{L} A_{ci} \xi_i^n + \sum_{i=1}^{M} A_{ci} \mu_i \xi_i^n \zeta_i^n \qquad (2.12)$$

Written in matrix form, Eqn. (2.12) becomes

$$\begin{bmatrix} y(0) \\ y(1) \\ y(2) \\ \vdots \\ y(N-1) \end{bmatrix} = \left[\begin{array}{ccc|ccc} 1 & \cdots & 1 & 1 & \cdots & 1 \\ \xi_1 & \cdots & \xi_L & \xi_1\zeta_1 & \cdots & \xi_M\zeta_M \\ \xi_1^2 & & \xi_L^2 & \xi_1^2\zeta_1^2 & & \xi_M^2\zeta_M^2 \\ \vdots & & \vdots & \vdots & & \vdots \\ \xi_1^{N-1} & \cdots & \xi_L^{N-1} & \xi_1^{N-1}\zeta_1^{N-1} & \cdots & \xi_{M1}^{N-1}\zeta_M^{N-1} \end{array}\right] \begin{bmatrix} A_{c1} \\ \vdots \\ A_{cL} \\ A_{c1}\mu_1 \\ \vdots \\ A_{cM}\mu_M \end{bmatrix} \qquad (2.13)$$

which should be sufficiently overdetermined with $N > L + M$. Now equation (2.13) is to be solved in the least square sense to find the A_{ci} and μ_i parameters. The pseudo inverse of the matrix here is computed by Singular Value Decomposition technique. Note that Eqn. (2.13) presents a linear estimation problem, and consequently it can be stated that the amplitude, phase and modulation index parameters will be computed accurately provided the angular frequency parameters are known accurately. [1]

After the estimation of all the parameters a data set is regenerated with these parameters, which is then compared with original input signal using Spectral Distortion measure [5]. A mismatch can occur between the regenerated and input signal because of wrong choice of unmodulating carrier and modulating frequencies. In such a case a new may have to be made and estimation process repeated till a proper match between the input and regenerated signal is obtained. [3]

2.3 Simulation Study

In this section, computer synthesized is considered for the study of model fitting.

The complex signal $y(n)$ consisting of two single tone-amplitude modulated signals and one unmodulated carrier signal is sampled at $N = 800$ points. Three sets of parameters in Eqn. (2.1) are chosen as follows:-

$$A_1 = 1.00,\ \omega_1 = 20,\ \nu_1 = 1,\ \phi_1 = 2,\ \mu_1 = 0.50,$$
$$A_2 = 0.50,\ \omega_2 = 40,\ \nu_2 = 0,\ \phi_2 = 3,\ \mu_2 = 0.00,$$
$$A_3 = 0.75,\ \omega_3 = 60,\ \nu_3 = 3,\ \phi_3 = 1.5,\ \mu_3 = 0.65,$$

and the sampling interval used is $T = 0.01$ units of time. Utilizing Eqn. (2.7) with $n_1 = J$ and $n_2 = N-1-J$, the AACF sequence $C_y(k)$ for $k = -J,\cdots,0,\cdots J$ is computed. It has been found empirically that for the best results value of J should be chosen as $N/3$ or $2N/5$ [1,3]. From the computed sequence of the AACF the AR parameters of the PEF are found

using modified covariance technique as given by Eqn. (2.9). Model order chosen is high enough (above 100) so that no frequency peak is missed out even in presence of high noise. Then the zeros of PEF are computed. Only those zeros, which are on or near the unit circle, are considered for further processing. As a first step the residues pertaining to the chosen frequencies (zeros in this case) are found. Using these residues we find the unmodulated carrier and modulating frequencies as discussed in section (2.2). Other parameters are determined using linear least square technique. True and estimated values of parameters along with percentage errors are listed in table (2.1). Residues for various frequencies are listed in table (2.2).

Table 2.1: Estimated values of complex AM model parameters for synthesized data

Parameters	True Value	Estimated values (percentage error)		
		No noise	SNR=30db	SNR=20db
ω_1	20.000	20.000(0.00)	20.001(0.00)	20.003(0.01)
ω_2	40.000	40.000(0.00)	40.002(0.01)	40.007(0.01)
ω_3	60.000	60.000(0.00)	59.999(0.00)	59.993(0.01)
v_1	1.000	1.000(0.00)	1.001(0.12)	1.024(2.40)
v_2	-	-	-	-
v_3	3.000	3.000(0.00)	3.001(0.02)	2.999(0.03)
A_1	1.000	1.000(0.00)	0.997(0.30)	0.979(2.10)
A_2	0.500	0.500(0.00)	0.498(0.40)	0.497(0.15)
A_3	0.750	0.750(0.00)	0.752(0.15)	0.753((0.26)
ϕ_1	2.000	2.000(0.00)	1.992(0.40)	1.975(1.25)
ϕ_2	3.000	3.000(0.00)	2.999(0.03)	2.959(1.36)
ϕ_3	1.500	1.500(0.00)	1.497(0.20)	1.485(1.00)
μ_1	0.500	0.500(0.00)	0.503(0.60)	0.521(4.2)
μ_2	-	-	-	-
μ_3	0.650	0.650(0.00)	0.652(0.13)	0.648((0.30)

Table 2.2: List of Frequencies and corresponding residues of complex AM model for synthesized data

(a): For 20db SNR

#	Freq.	Residue
1	63.992	53.06 + j8.90
2	59.993	96.11 - j9.67
3	40.007	58.40 - j3.41
4	21.027	-8.40 + j52.74
5	20.003	109.83 –
6	16.58	46.97
7	268.4	0.19 + j0.04
		0.10 + j0.01

(b): For 30db SNR

#	Freq.	Residue
1	20.001	110.37 –j50.98
2	21.002	-7.72 – j55.65
3	40.002	58.02 – j3.60
4	96.99	0
5	59.999	103.91 – 11.22
6	63.000	51.97 +j10.15

(c) for zero noise

#	Freq.	Residue
1	21.00	110.33 – 50.82
2	20.00	-7.65 – j55.32
3	40.00	58.59 – j3.67
4	52.84	0
5	60.00	103.87 –j10.93
6	63.000	51.92+ j10.09

Chapter 3

Complex FM Signal Model

3.1 The Model

It has been shown [2] that Complex FM Signal Model can be used for representation of non-stationary signals like unvoiced speech phonemes. The complex random sequence $x(n)$ consisting of M single tone frequency modulated subsequences is represented by

$$x(n) = \sum_{i=1}^{M} A_i \, e^{j[\omega_i nT + \phi_i + \beta_i \sin(v_i nT)]} \tag{3.1}$$

where

A_i is the amplitude of complex exponential carrier signal,

ω_i is the discrete carrier angular frequency,

β_i is the modulation index for sinusoidal modulating signal,

v_i is the discrete modulating angular frequency,

T is the sampling interval,

ϕ_i is the random phase, independent and identically distributed.

The random phase ϕ_i is assumed to be uniformly distributed over $[0,2\pi]$. It has been shown [2] that unlike Complex AM Signal Model case,

Time-Varying Autocorrelaion Function (ACF) can not be conveniently utilized in this case, for determining the carrier and modulating angular frequencies because of its double dependence on time and lag.

Instead another function called Product Function has been utilized for this purpose [3]. The product function $p_x(k)$ is defined as

$$p_x(k) = \left[x^*(n)x(n+k)\right]_{n=k/2};\ k = 0, \pm 2, \pm 4, \cdots \qquad (3.2)$$

evaluating the expression (3.2) for sum of FM signals $x(n)$, one gets

$$p_x(k) = \sum_{i=1}^{M}\sum_{l=1}^{M} A_i A_l e^{j(\omega_i+\omega_l)Tk/2} . e^{j(\phi_i-\phi_l)} . e^{j[\beta_i \sin(v_i Tk/2)+\beta_l \sin(v_l Tk/2)]}$$

In order to investigate the characteristics of product function $p_x(k)$, we consider a specific case of M=2. Then the above expression for $p_x(k)$ after simplification [2,3] can be written as follows: -

$$p_x(k) = A_1^2 \sum_{m_1=-\infty}^{\infty} J_m(2\beta_1)\, e^{j(2\omega_1+mv_1)k/2}$$

$$+ A_2^2 \sum_{m_2=-\infty}^{\infty} J_m\left(2\beta_2\right) e^{j(2\omega_2+mv_2)k/2}$$

$$+ A_{12} \sum_{m_1=-\infty}^{\infty}\ \sum_{m_2=-\infty}^{\infty} J_{m_1}\left(2\beta_1\right) J\left(2\beta_2\right) e^{j(\omega_1+\omega_2+m_1v_1+m_2v_2)k/2}$$

where $A_{12} = A_1 A_2\, e^{j(\phi_2-\phi_1)} + A_2 A_1\, e^{j(\phi_1-\phi_2)}$

It is clear from the equation that the spectrum of the sequence $p_x(k)$ will contain peaks at twice the carrier angular frequency and the sum of

carrier angular frequencies, with an infinite number of side peaks located symmetrically on either side of the main peaks. The clusters centered at twice the carrier frequencies are like individual FM signals.

3.2 Model Fitting and Parameter Estimation

For the sum of complex FM signals $x(n)$, a sequence of product function $p_x(k)$ is computed. This sequence is then fitted into a linear prediction model and zeros of the PEF [4] are computed as in the complex AM signal model case. Theses zeros indicate the frequencies contained in the signal and are processed further to obtain carrier and modulating frequencies of the model.

As brought out in section (3.1), the product function contains a very large number of frequencies. Therefore, the AR model order in this case has to be chosen as a very high value (greater than 150) to ensure that significant peaks are not missed out. Identifying and separating carrier angular and modulating frequency is a very difficult task. For this, frequencies obtained by the zeros of PEF, are sorted in ascending order and difference (subtracting lower frequency from just higher one) of two consecutive frequencies is tabulated. By this table a cluster can be identified, and with the knowledge that each cluster is centered at twice the

carrier frequency and both sides of this center are frequencies which are sum of twice the carrier frequency and multiples of modulating frequency, carrier and modulating frequencies can be determined from their respective clusters.

The sum of carrier frequencies also makes a cluster and this can be confused with unmodulated carrier frequency so whenever this type of conflict arises, hit and trial method is used.

DFT plot of data $x(n)$ may also have to be used as an aid for accomplishing the above task. Isolated peaks in the spectrum are defined as unmodulated carrier frequencies.

Once the carrier and modulating frequencies have been determined next problem is to determine respective modulation indices and other parameters. We can write Eqn. (3.1) as

$$x(n) = Ac_1 \sum_{m=-\infty}^{\infty} J_m(\beta_1) e^{j(\omega_1 + m\nu_1)nT} + \cdots + Ac_M \sum_{m=-\infty}^{\infty} J_m(\beta_M) e^{j(\omega_M + m\nu_M)nT}$$

where $Ac_i = A_i e^{j\phi_i}$ is the complex amplitude and $J_m(\beta_i)$ is the Bessel Function of the first kind of integer order m and argument β_i [10].

It can be seen from above equation that each of the FM subsignals of $x(n)$ with non-zero β_i would contain an infinite number of side frequencies. But it is known [9] that for a FM signal with as high as 30; the number of significant side frequencies is just 70. Therefore we take a fixed number of side frequencies (say q) and determine their residues as follows: -

$$\begin{bmatrix} x(1) \\ x(2) \\ \vdots \\ x(N) \end{bmatrix} = \begin{bmatrix} e^{j(\omega_1 - qv_1)} & e^{j2(\omega_1 - qv_1)} & \cdots & e^{jN(\omega_1 - qv_1)} \\ \vdots & \vdots & \cdots & \vdots \\ e^{j\omega_1} & e^{j2\omega_1} & \cdots & e^{jN\omega_{11}} \\ \vdots & \vdots & \cdots & \vdots \\ e^{j(\omega_1 + qv_1)} & e^{j2(\omega_1 + qv_1)} & \cdots & e^{jN(\omega + qv_1)} \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ e^{j(\omega_M - qv_M)} & e^{j2(\omega_M - qv_M)} & \cdots & e^{jN(\omega_M - qv_M)} \\ \vdots & \vdots & \cdots & \vdots \\ e^{j\omega_M} & e^{j2\omega_M} & \cdots & e^{jN\omega_m} \\ \vdots & \vdots & \cdots & \vdots \\ e^{j(\omega_M - qv_M)} & e^{j2(\omega_M - qv_M)} & \cdots & e^{jN(\omega_M - qv_M)} \end{bmatrix}^T \begin{bmatrix} Ac_1 J_{-q}(\beta_1) \\ \vdots \\ Ac_1 J_0(\beta_1) \\ \vdots \\ Ac_1 J_q(\beta_1) \\ \vdots \\ \vdots \\ Ac_M J_{-q}(\beta_M) \\ \vdots \\ Ac_M J_0(\beta_M) \\ \vdots \\ Ac_M J_q(\beta_M) \end{bmatrix} \quad (3.4)$$

where q is suitably chosen so that above set of equations remains overdetermined.

$\left[Ac_i J_m(\beta_i)\right]_{i=1,2,\cdots,M \text{ and } m=-q,\cdots,-1,0,1,\cdots,q}$ are the residues for respective frequencies of the signal $x(n)$.It is known [10] that

$$\sum_{m=-\infty}^{\infty} J_m^2(\beta) = 1 \quad (3.5)$$

Therefore sum of squares of residues for each FM subsignal gives us complex amplitude for that signal as follows: -

$$\left[Ac_i J_{-q}(\beta_i)\right]^2 + \cdots + \left[Ac_i J_0(\beta_i)\right]^2 + \cdots + \left[Ac_i J_q(\beta_i)\right]^2 \approx Ac_i^2 \quad (3.6)$$

Knowing complex amplitude, we can determine $J_m(\beta_i)$ from residues and hence the value of modulation index β_i can be determined for each FM subsignal. It is to be noted that while performing squaring operation as indicated above, the phase information of the complex

amplitude is lost. Hence we try for both positive and negative and negative values of complex amplitude and regenerate the residues after finding modulation index for each case. The values of complex amplitude and modulation index which regenerate the original residues are chosen.

After estimation of all the parameters a data set is regenerated with theses parameters, which is then compared with the original input signal using Spectral Distortion Measure [5]. In case of mismatch, a new choice may to be made and estimation process repeated till a proper match between the input and regenerated signals is obtained.

3.3 Simulation Study

In this section, computer synthesized data is considered for model fitting by Complex FM signal model.

The complex sequence $x(n)$; $n = 0, \pm 1, \cdots, \pm N/2$ consisting of two single-tone frequency modulated and one unmodulated carrier subsignals is sampled at $N = 801$ points. The sets of parameters of model (3.1) are chosen as follows: -

$$A_1 = 1.50,\ \omega_1 = 60,\ v_1 = 2,\ \phi_1 = 1.5,\ \beta_1 = 0.40$$
$$A_2 = 2.25,\ \omega_2 = 80,\ v_2 = 0,\ \phi_2 = 2.25,\ \beta_2 = 0.00$$
$$A_3 = 0.75,\ \omega_3 = 140,\ v_3 = 7,\ \phi_3 = 1.00,\ \beta_3 = 1.00$$

and the sampling interval $T = 0.01$ units of time .

Utilizing the sampled sequence, product function sequence $\{p_x(k)\}$ is computed using Eqn. (3.2) for $k = 0, \pm 2, \pm 4, \cdots, \pm L$ with $L = 800$ the 801 point sequence is then fitted into a linear prediction model and AR coefficients are determined using modified covariance technique as discussed in chapter 2. Zeros of PEF are the various frequencies contained in the product function sequence. Then theses frequencies are arranged in ascending order and clusters are selected, as explained earlier each cluster will double carrier frequency at its center. Help of PSD plot of data is also taken to identify the unmodulated carrier frequency. After estimation of carrier and modulating frequency, other parameters modulation index, amplitude and phase are found out using Eqn. (3.4) and (3.6). The value of q is chosen to be 5. $x(n)$ is then mixed with varying degree of noise and estimation procedure is repeated. True and estimated values of various parameters along with the percentage errors are listed in table (3.1).

Frequencies and difference frequencies are also shown to explain the identification of cluster and finding carrier and modulating frequencies.

Table 3.1: Estimated values of complex FM model parameters for synthesized data

Parameters	True Value	Estimated values (percentage error)		
		No noise	SNR=30db	SNR=20db
ω_1	60.000	60.000(0.00)	60.005(0.00)	60.000(0.00)
ω_2	80.000	80.000(0.00)	80.000(0.01)	80.000(0.00)
ω_3	140.00 0	140.000(0.00)	140.000(0.00)	140.000(0.00)
v_1	2.000	2.000(0.00)	2.001(0.12)	2.12(6.00)
v_2	-	-	-	-
v_3	7.000	7.000(0.00)	7.000(0.00)	7.000(0.00)
A_1	1.500	1.500(0.00)	1.501(0.07)	1.445(3.60)
A_2	1.250	1.250(0.00)	1.252(0.15)	1.254(0.32)
A_3	0.750	0.750(0.00)	0.749(0.13)	0.747((0.40)
ϕ_1	1.500	1.500(0.00)	1.497(0.22)	1.496(0.26)
ϕ_2	2.250	2.250(0.00)	2.252(0.07)	2.254(0.18)
ϕ_3	1.000	1.000(0.00)	1.005(0.48)	1.015(1.51)
β_1	0.400	0.400(0.00)	0.40(0.00)	0.43(7.5)
β_2	-	-	-	-
β_3	1.000	1.000(0.00)	1.000(0.00)	1.01(1.00)

Chapter 4

Speech Coding

Parametric modeling by complex AM and FM signal models has been studied for sustained vowel and consonant sounds [3]. But in natural speech the signal is not steady i.e. the energy or gain of the speech varies with time. Hence the proposed models can not be applied for continuously spoken speech with varying gain. In this section the time varying nature of speech gain is considered and before applying these models speech signal is made of constant gain.

4.1 Estimation and Modeling of Gain

Let $s(n)$ be the speech sequence with varying gain, consider another sequence $y(n)$ such that,

$$y(n) = \frac{s(n)}{g(n)} \qquad (4.1)$$

where $g(n) > 0, \forall n$.

The quantity $g(n)$ is the time varying gain of speech sequence $s(n)$. Now the resulting sequence $y(n)$ is of constant gain or variance. It remains to show how one gets $y(n)$ from $s(n)$, in other words, how the gain factor $g(n)$ is

estimated for each n. As it is already assumed that $g(n)$ evolves slowly with respect to time, one can estimate it as the local envelope or short time energy of $s(n)$. [11]

For a symmetric window 'w' of even length L, $g(n)$ can be estimated as,

$$\hat{g}(n) = \frac{1}{L}\sum_{i=1}^{L}|w(i)s(n+i-L/2)| \qquad (4.2)$$

for $n = \frac{L}{2},\cdots,(N-\frac{L}{2})$.

The choice of window length L depends on the nature of the signal $s(n)$ and that of the sequence $g(n)$. For very small L, the averaging is not complete and the randomness of $s(n)$ is reflected in $\hat{g}(n)$. On the other hand if L is too large, it fails to follow the local variation of the envelope.

Here we see that the length of the gain sequence $\hat{g}(n)$ is $N-L$, while that of sequence $s(n)$ is N. Due to this problem the division of $s(n)$ by $\hat{g}(n)$ is not possible. This difficulty is solved by adding zeros of length $L/2$ on both sides of sequence $s(n)$.

Now the gain sequence $\hat{g}(n)$ is fitted by polynomial model as follows: -

$$\hat{g}(n) = p_1 n^m + p_2 n^{m-1} + p_3 n^{m-2} + ,\cdots, p_m n + p_{m+1} \qquad (4.5)$$

where m is the model order and p_i's are the polynomial coefficients.

4.2 Estimation of Speech Parameters

Before estimating the parameters one should know the time interval of each phoneme of continuously spoken speech. This phoneme separation is not the matter of this study, here it is assumed that we are already having various separated phonemes.

Now each phoneme sequence is selected and gain is estimated for the total period of this phoneme sequence. After estimation of gain, phoneme sequence is divided by this gain for whole of its length. By this way we are now having constant gain phoneme sequence and complex AM or FM signal models can be applied on this.

. It has been shown in [3] that complex AM signal model is suitable for vowel sounds while FM signal model is suitable for consonant sounds. But simulation studies show that this type of classification is not appropriate. Many consonant sounds can also be modeled by complex AM signal model. It is found by simulation studies that the phonemes having most of the energy in low frequency region can be modeled by complex AM signal model and the phonemes having most of the energy in high frequency region can be modeled by complex FM signal model.

Frequency range of phoneme is seen by simple PSD plot. Though we know that the standard Fourier representations that are appropriate for periodic, transient, or stationary random signals are not directly applicable to the representation of the speech signal whose properties change markedly as a function of time. Yet to have just an idea of the frequency content of *phoneme simple PSD plot of the phoneme is taken. After knowing which* model to apply for each phoneme every phoneme can be modeled either by complex AM or by complex FM signal model as discussed in chapter 2 and 3.

4.3 Simulation Study

In this section natural speech signal, used in TIMIT lexicon and phonetic transcription is taken. A sentence "don't ask me to"spoken by a women is chosen for model fitting and estimation of parameters. The phonetic details about the sentence are given in table (4.1) and (4.2). Speech data taken is sampled at 16 kHz. The signal sequence for each phoneme is normalized and made zero mean before model fitting.

For each phoneme *PSD plot is taken and complex AM or FM signal* model is applied according to section 4.2. All the 11 phonemes given in table 4.2 could be faithfully regenerated after fitting them into the complex AM or

FM model and estimating the parameters. The original and regenerated signals together with various intermediate steps are explained for phonemes 'n', 'dx', 'ix' and 's'.

For phoneme 'n' the PSD plot is shown in figure (4.6) .by seeing the figure it is clear that most of the energy is in low frequency region so it can be fitted by complex AM model. The gain for this phoneme signal is estimated by a Hanning window of length $L = 140$. The gain sequence is fitted by a polynomial of order 4. Coefficients of the polynomial are given in table (4.4). The original and regenerated gain is shown in figures (4.3) and (4.4) respectively. The original signal of phoneme is divided by gain sequence and resultant signal is shown in figure (4.5(a)). This signal obtained after division of gain is fitted by complex AM model. And all parameters are determined as discussed in chapter 2. After estimation of parameters signal is regenerated and shown in figure (4.5(b)).

Finally the regenerated signal shown in figure (4.5(b)) is multiplied with regenerated gain shown in figure (4.4), to get original phoneme signal of varying gain as shown in figure (4.1).

Various estimated parameter values for the regenerated signal for phoneme 'n' are shown in table (4.3).

The PSD plots for the phonemes 'dx' and 'ix' are shown in figures

(4.12) and (4.18) respectively . It is apparent from the plot that these phonemes can also be fitted by complex AM signal model. The Hanning window length for estimation of gain of phoneme 'dx' is $L = 140$ while for phoneme 'ix' length is taken to be 150.The gain sequence of phoneme 'dx' is fitted by polynomial of model order 5, while that for 'ix' model order chosen is 6. Various parameters for these two phonemes are estimated in the same way as for phoneme 'n'. Final original and regenerated signals and various intermediate figures for phoneme 'dx' are plotted in figures (4.7),(4.8) ,(4.9),(4.10),and (4.11)

Final and regenerated signals and various intermediate figures for phoneme 'ix' are plotted in figures (4.13), (4.14), (4.15), (4.16), and (4.17)

Estimated parameters of phoneme 'dx' and 'ix' are shown in table (4.5) and (4.7) respectively. The coefficients of polynomial for phoneme 'dx' and 'ix' are shown in tables (4.6) and (4.8) respectively.

From the PSD plot of phoneme 's' as shown in figure (4.24) ,it is clear that most of the energy lies in high frequency region . To fit this signal complex FM signal model is used. The gain is estimated with Hanning window of length $L = 300$, and fitted by polynomial of model order 8. The product function sequence is then fitted into a linear prediction model. The AR model order chosen is greater than 180. Using the zeros of PEF, and making clusters as explained in section (3.2), frequency parameters of

the model are determined. Then using these frequencies all other parameters of the model are determined employing the method discussed in section (3.2).

Final regenerated signal and original signal with varying gain for this phoneme are shown in figures (4.19) and (4.20) respectively. Various other intermediate plots are shown in figures (4.21), (4.22), and (4.23)

Estimated parameters of the phoneme and polynomial coefficients for the gain are listed in tables (4.9) and (4.10)

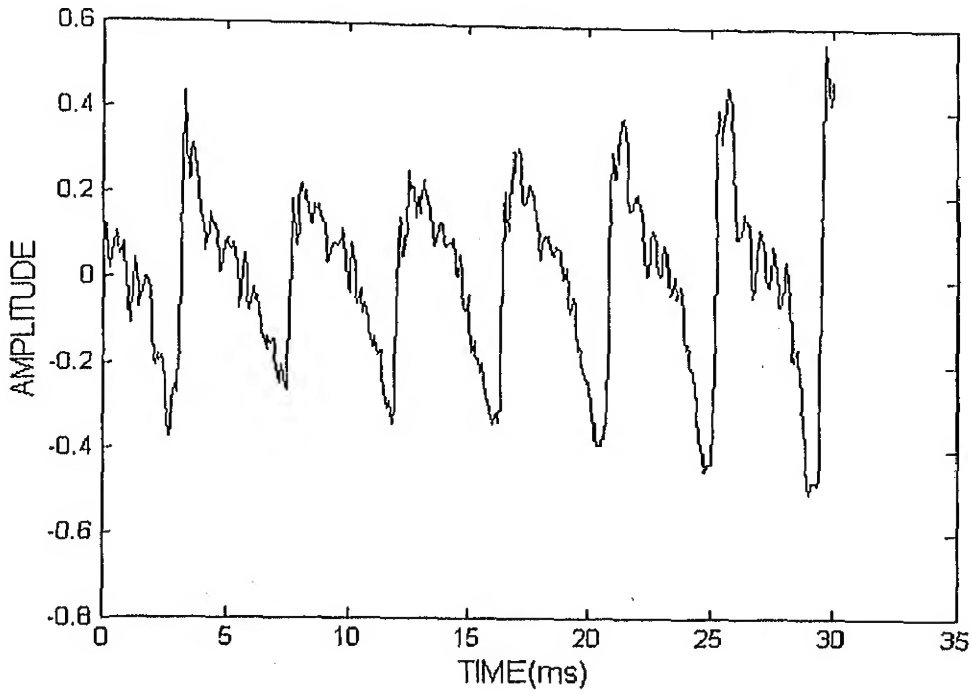


Fig.4.1: The original phoneme sound 'n'

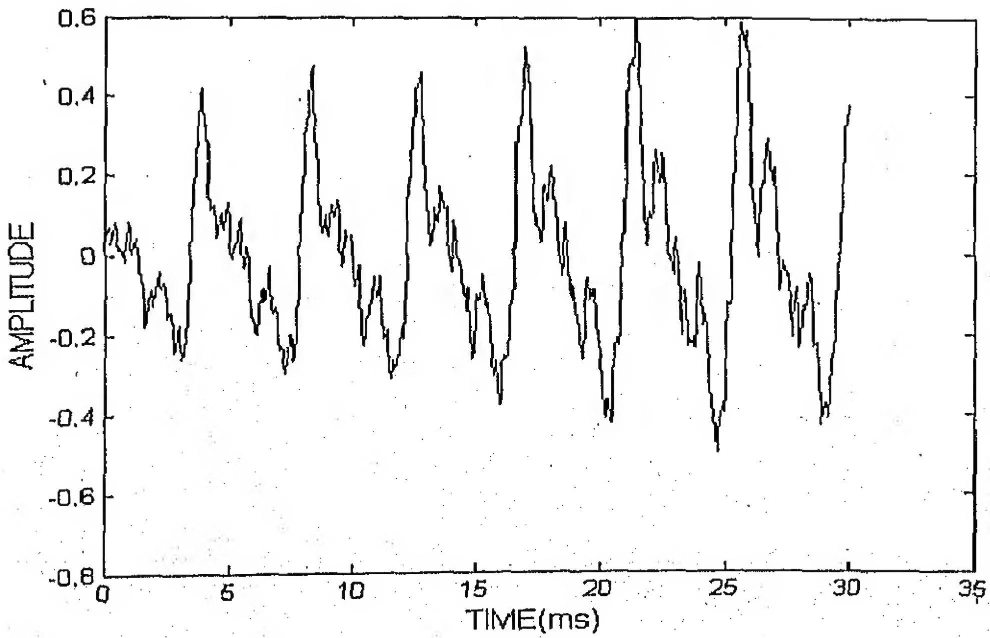


Fig.4.2: The regenerated phoneme sound 'n'

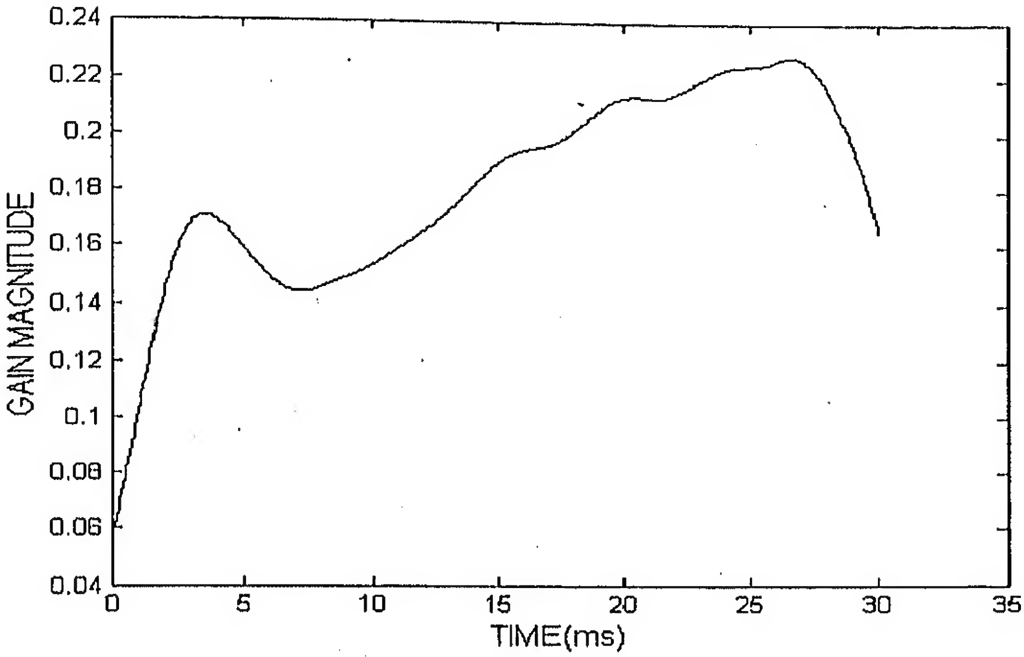


Fig.4.3: The Original gain function of phoneme sound 'n'

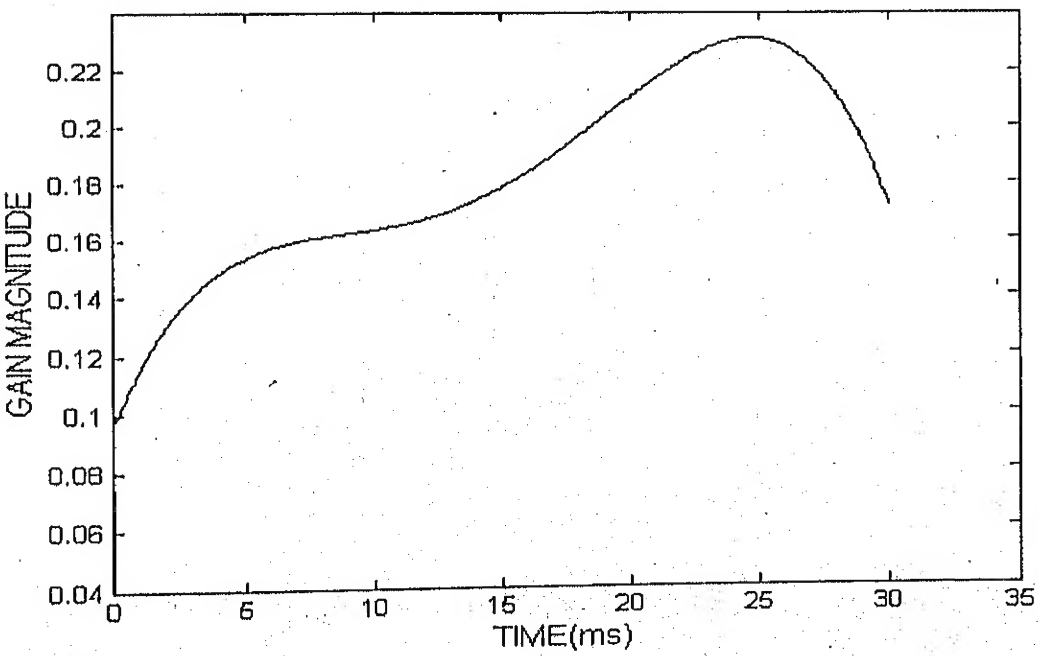
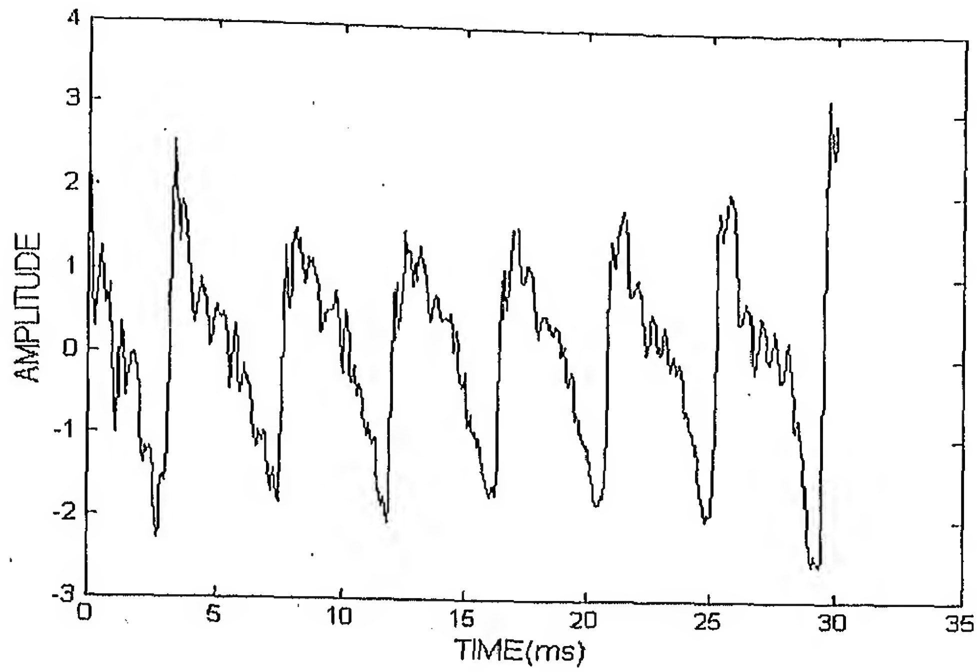
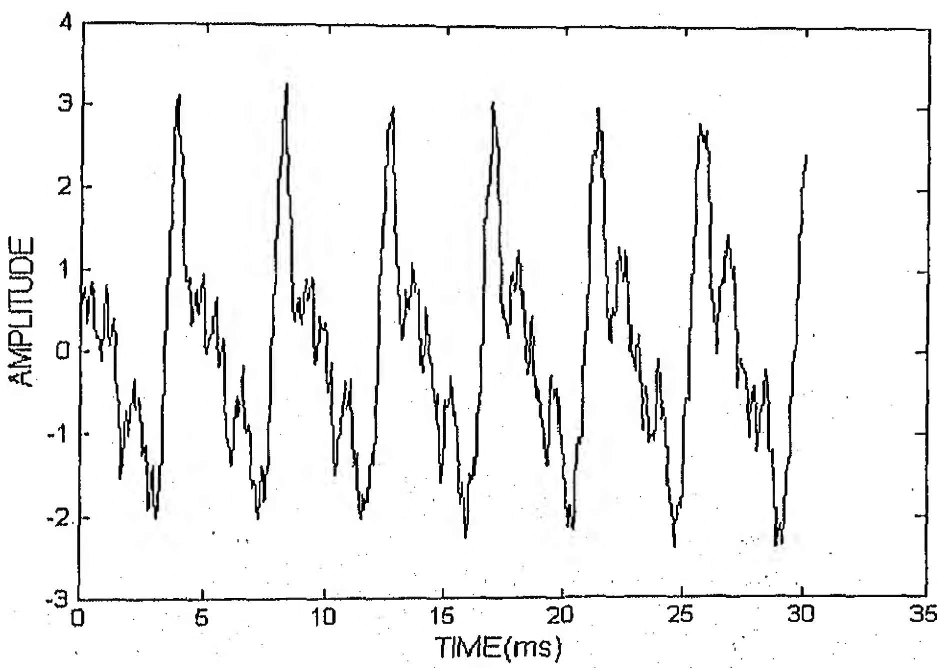


Fig.4.4: The regenerated gain function of phoneme sound 'n'



(a)



(b)

Fig.4.5: The phoneme sound n after division of gain (a) Original
(b) Regenerated

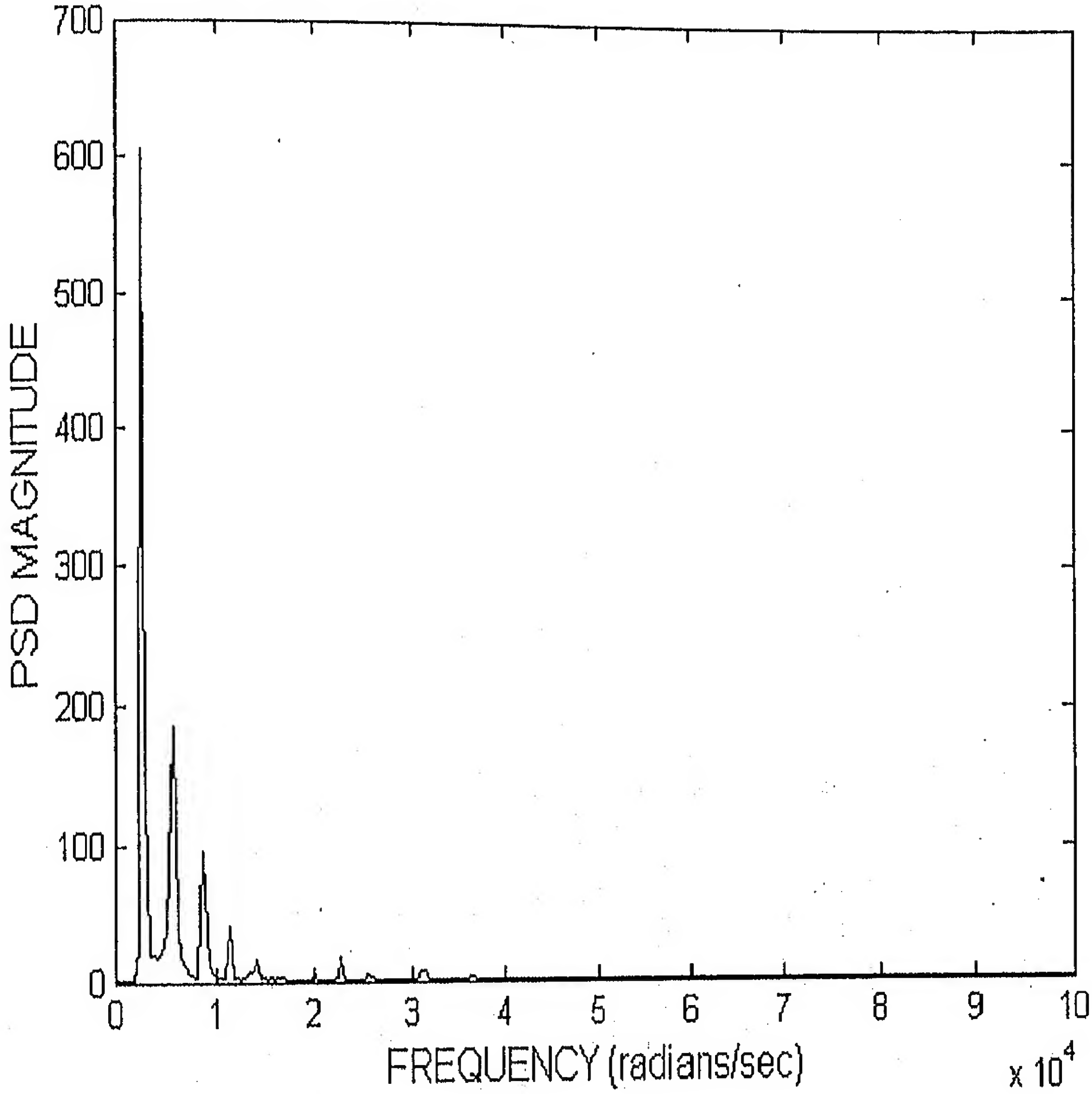


Fig.4.6: The PSD plot of phoneme sound 'n'

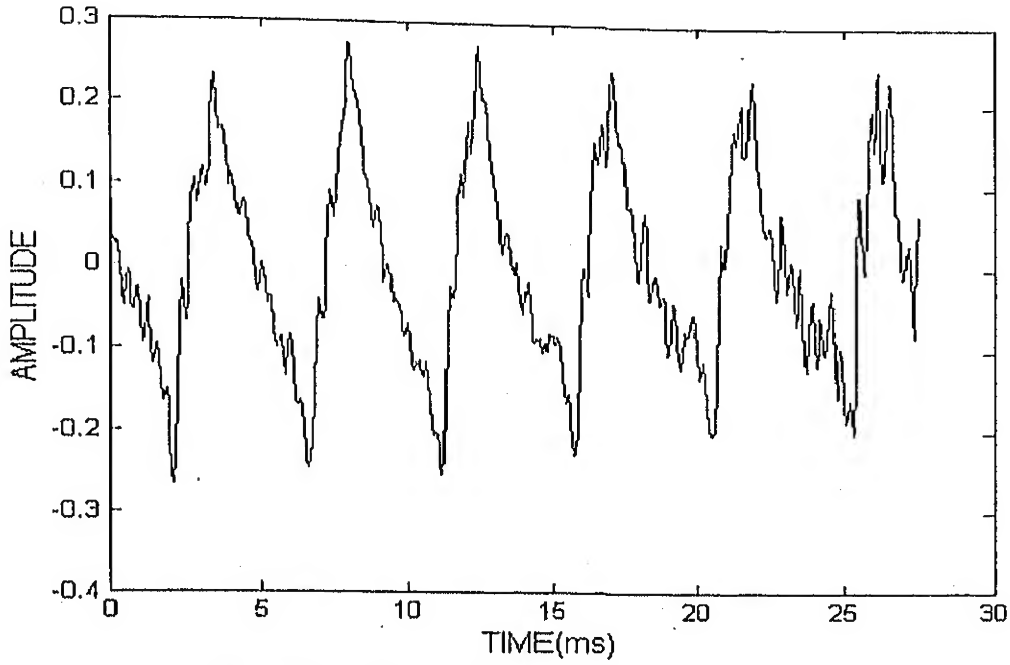


Fig.4.7: The original phoneme sound 'dx'

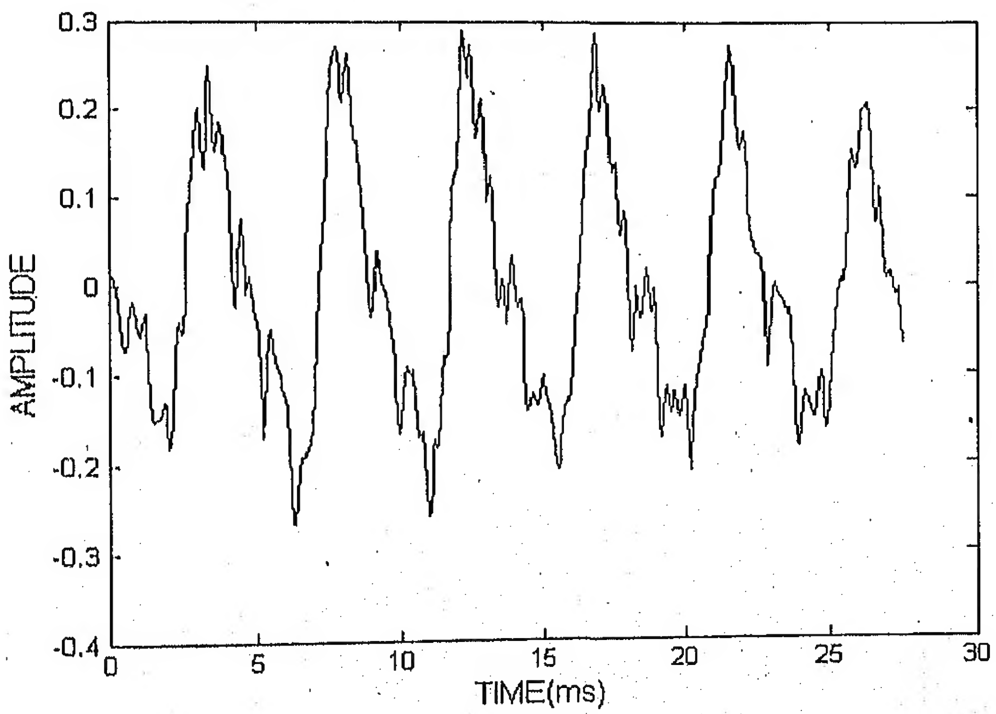


Fig.4.8: The regenerated phoneme sound 'dx'

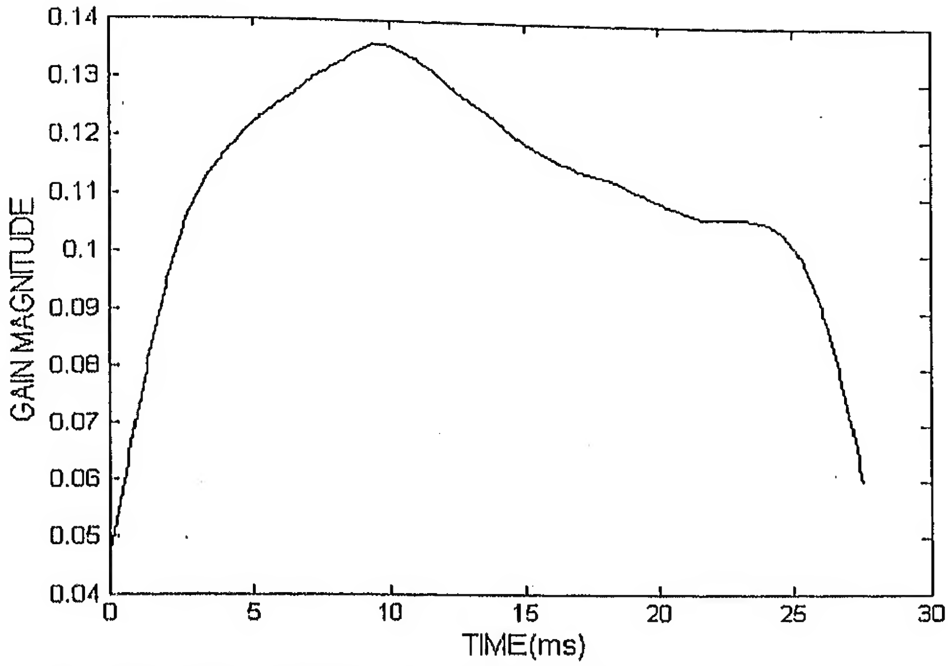


Fig. 4.9: The original Gain function function of phoneme sound 'dx'

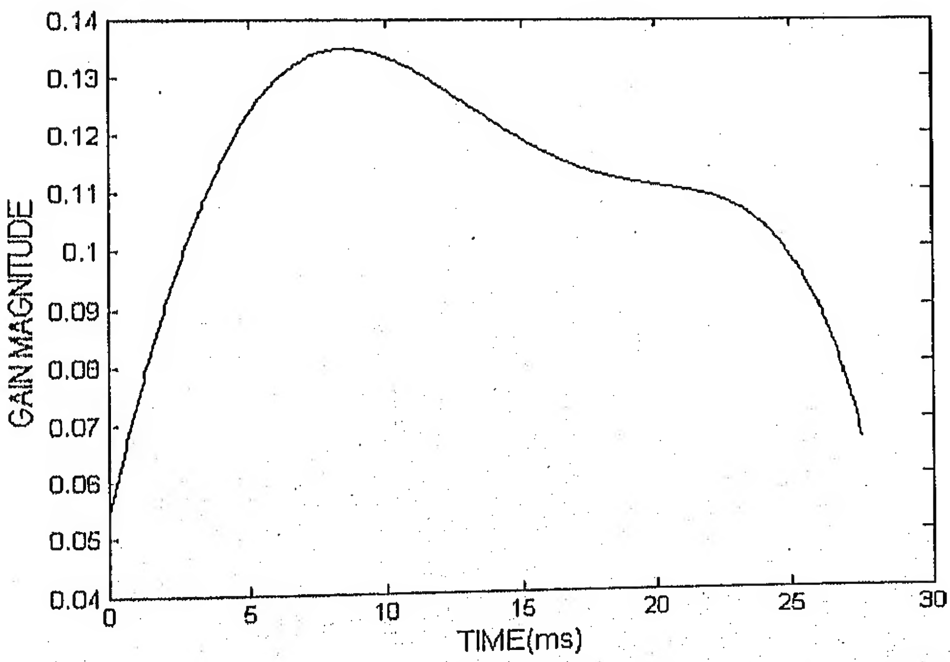
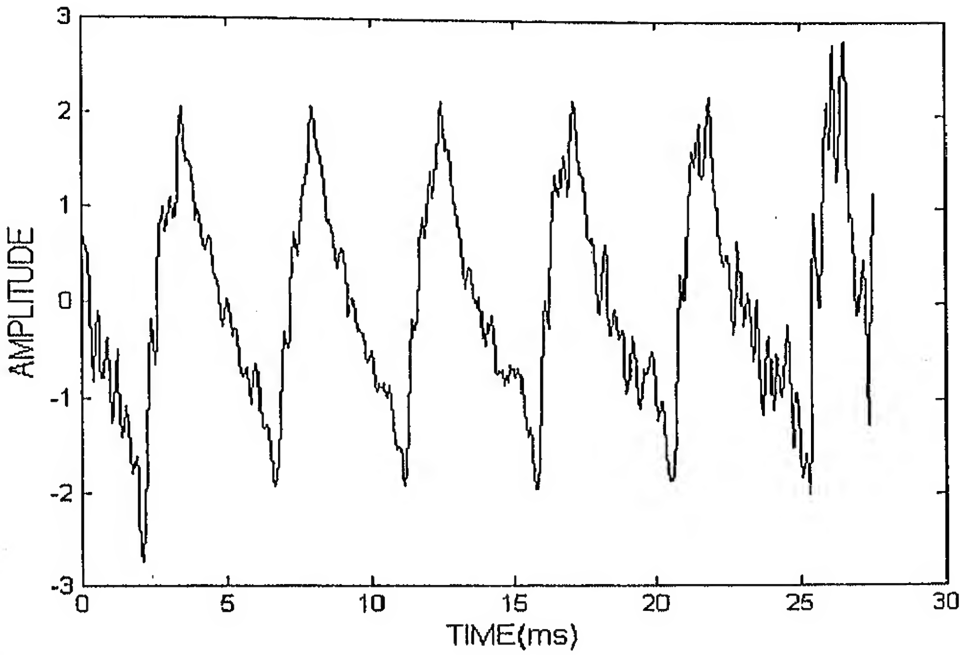
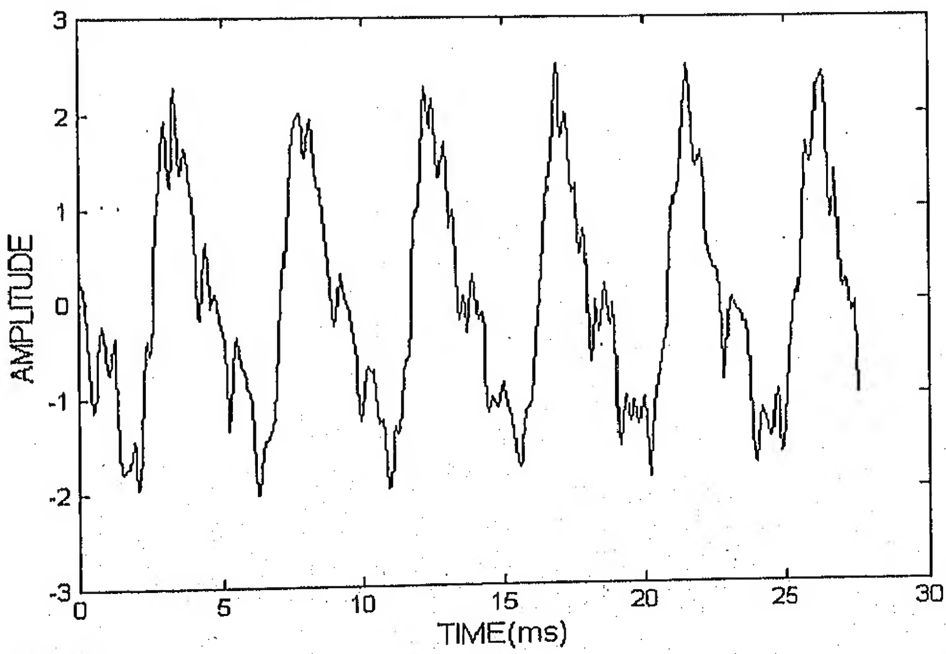


Fig.4.10 The regenerated Gain function of phoneme sound 'dx'



(a)



(b)

Fig .4.11 The phoneme sound 'dx' after division of gain function (a)Original (b) Regenerated

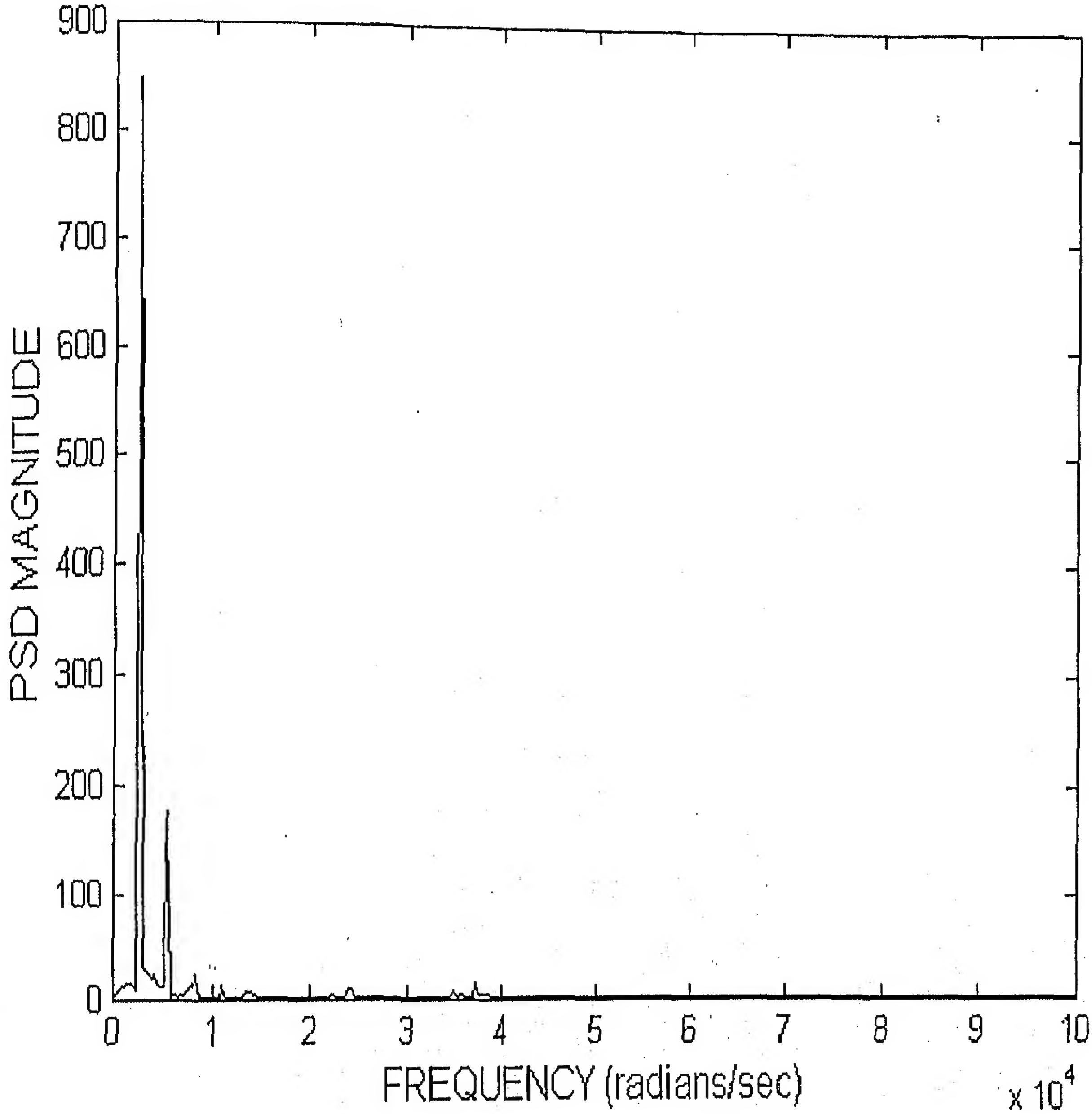


Fig.4.12: The PSD plot of phoneme sound 'dx'

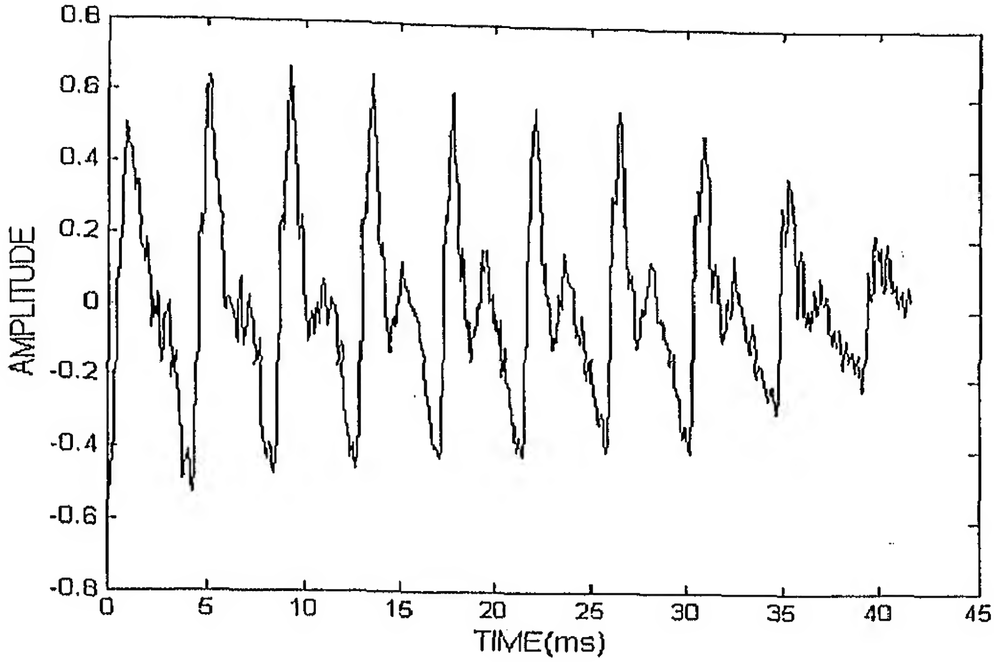


Fig.4.13: The original phoneme sound 'ix'

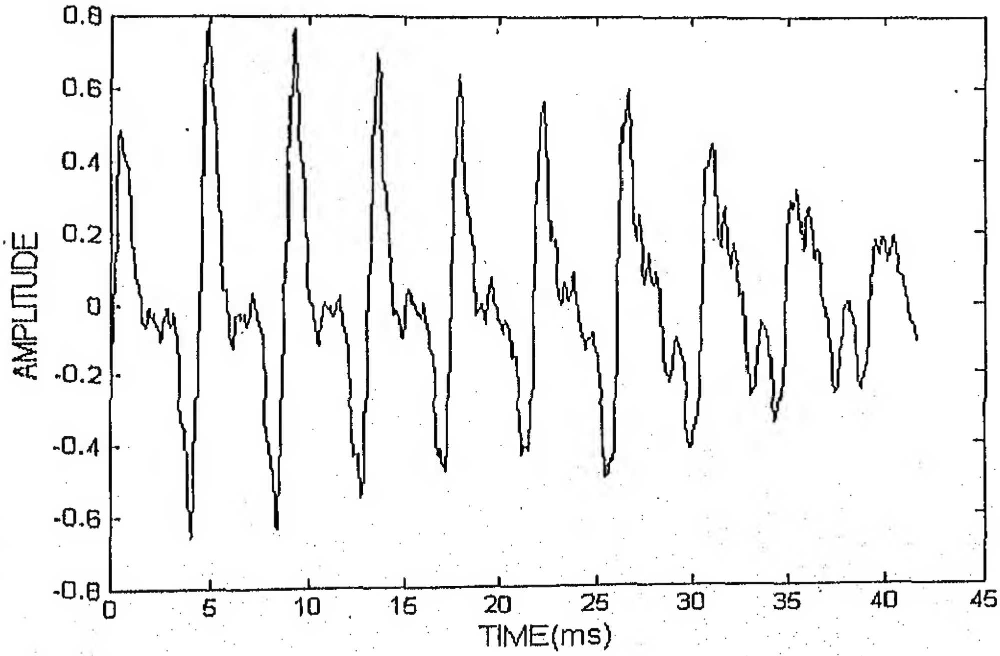


Fig.4.14: The regenerated phoneme sound 'ix'

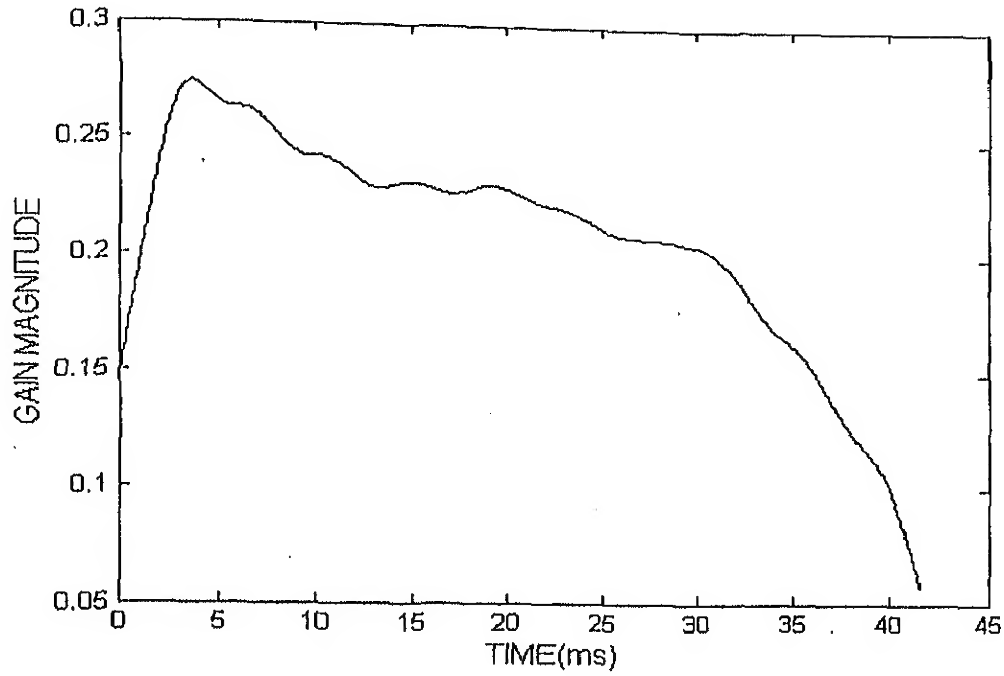


Fig4.15: The original Gain function of phoneme sound 'ix'

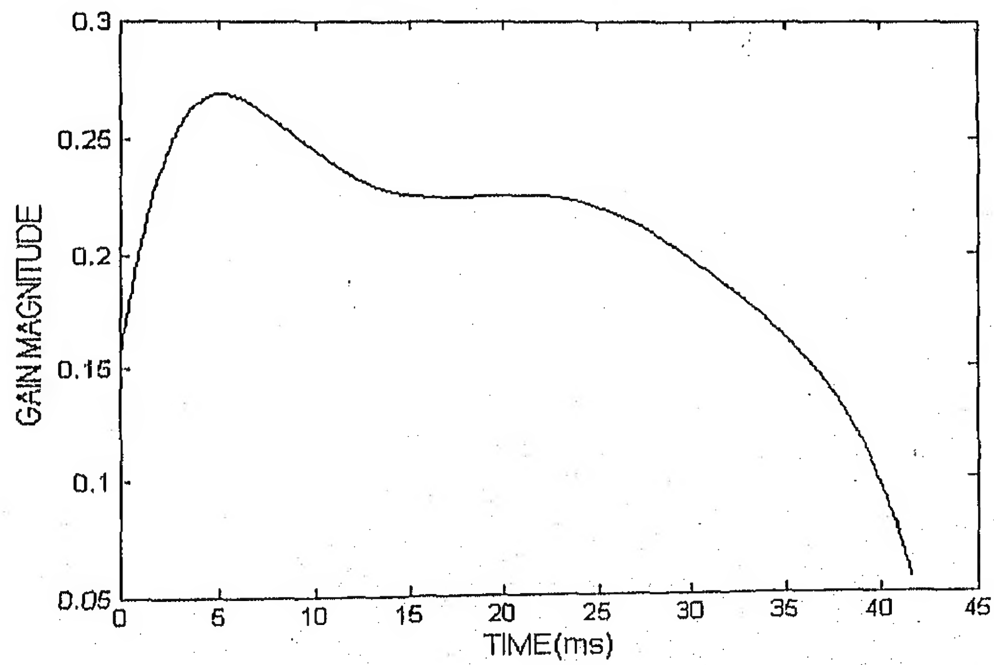
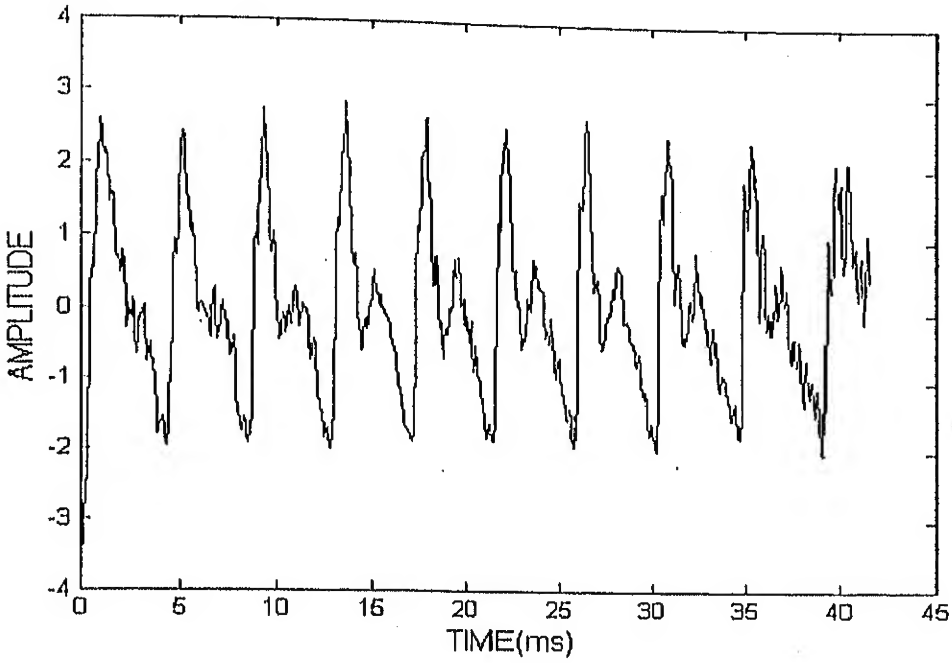
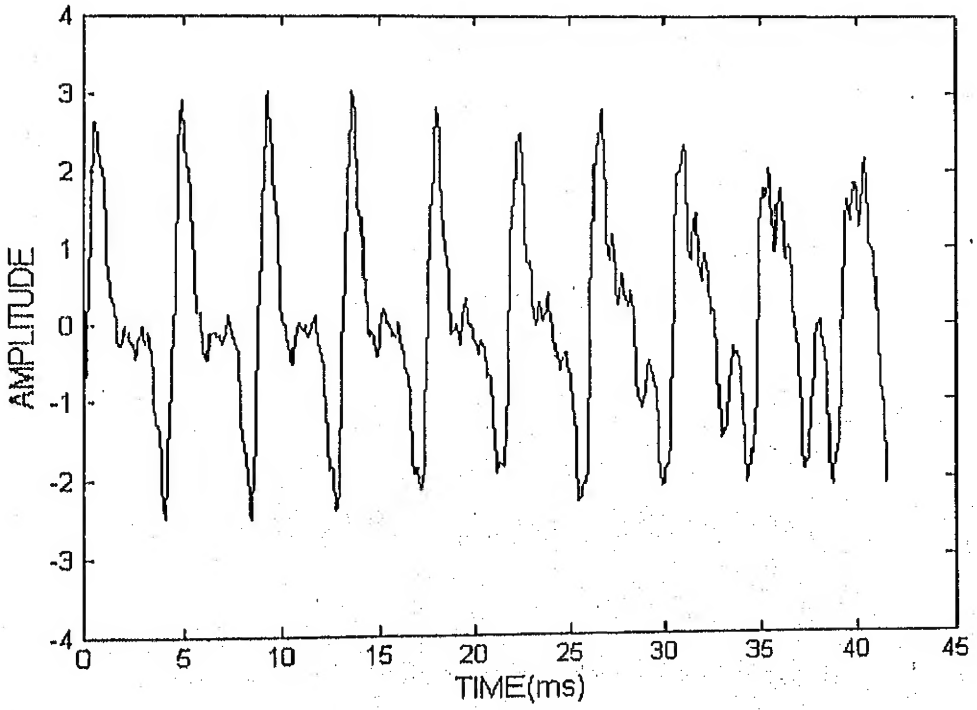


Fig.4.16: The regenerated Gain function of phoneme sound 'ix'



(a)



(b)

Fig 4.17: The phoneme sound 'ix' after division of gain (a) Original
(b) Regenerated

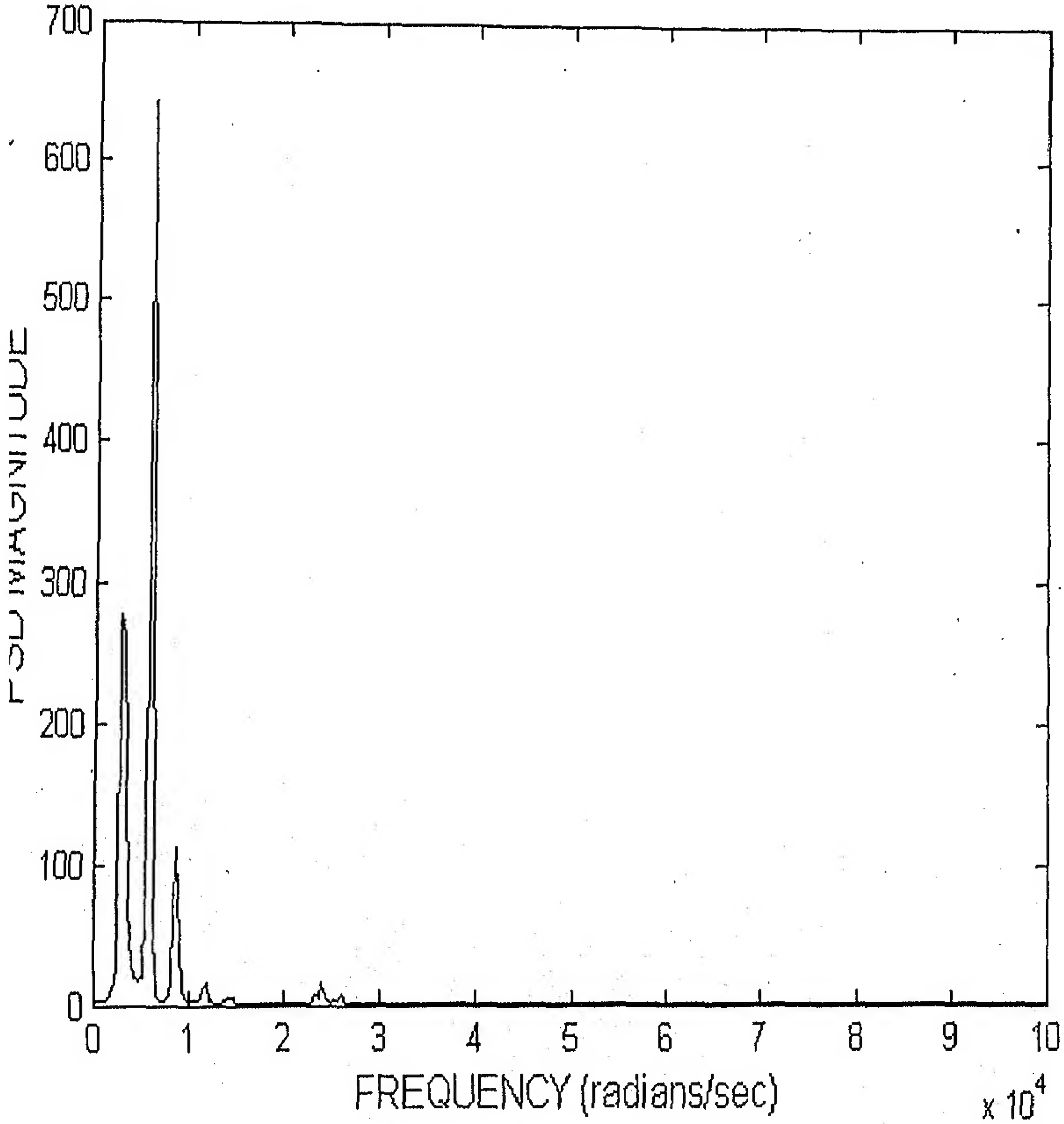


Fig.4.18: The PSD plot of phoneme sound 'ix'

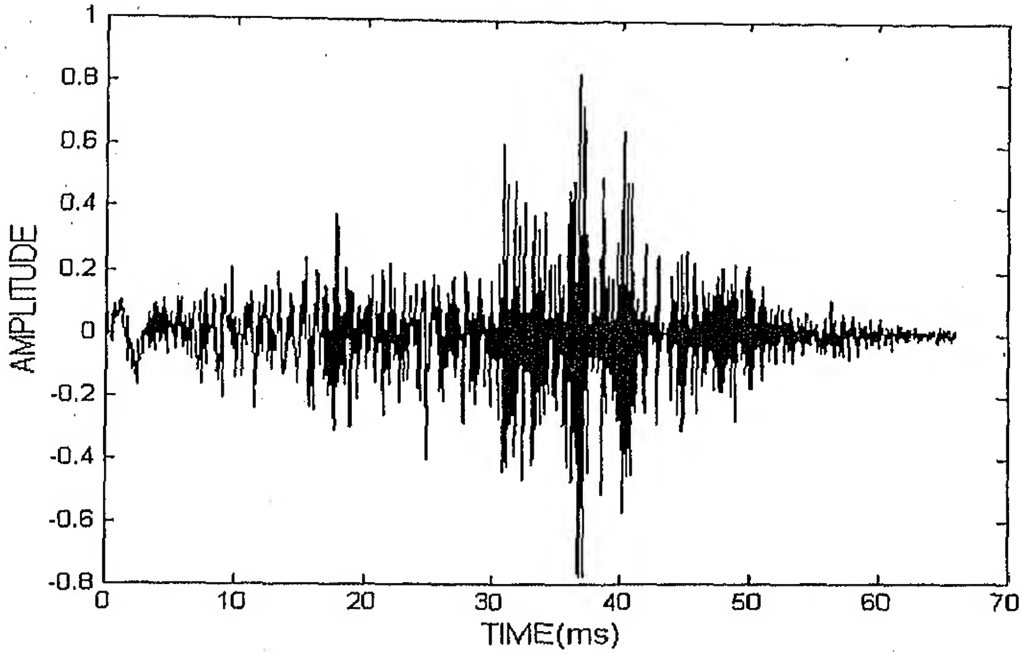


Fig.4.19: The original phoneme sound 's'

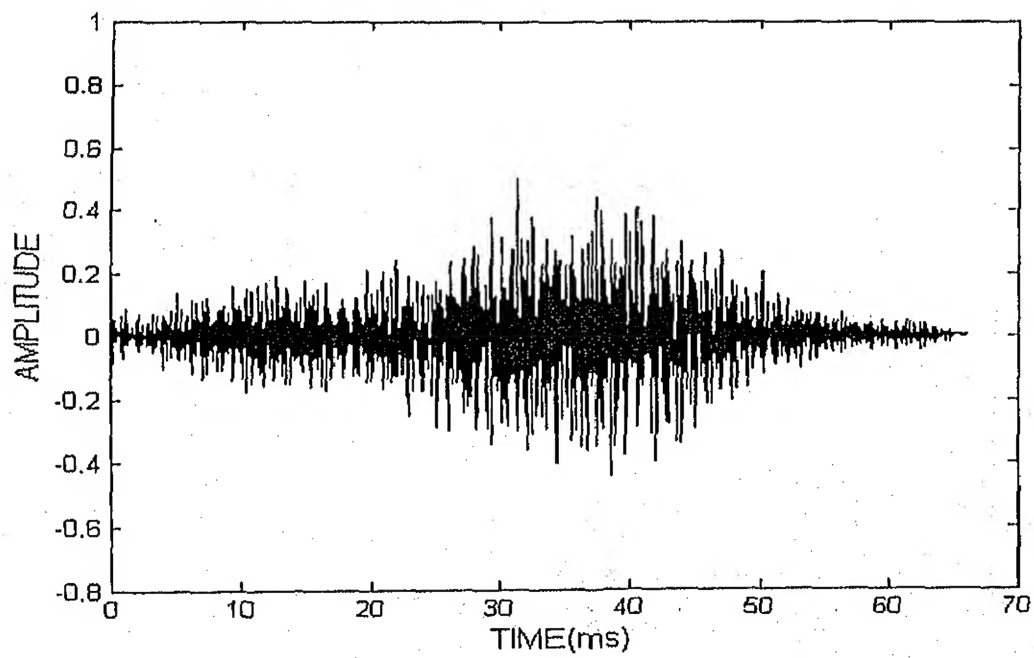


Fig.4.20: The regenerated phoneme sound 's'

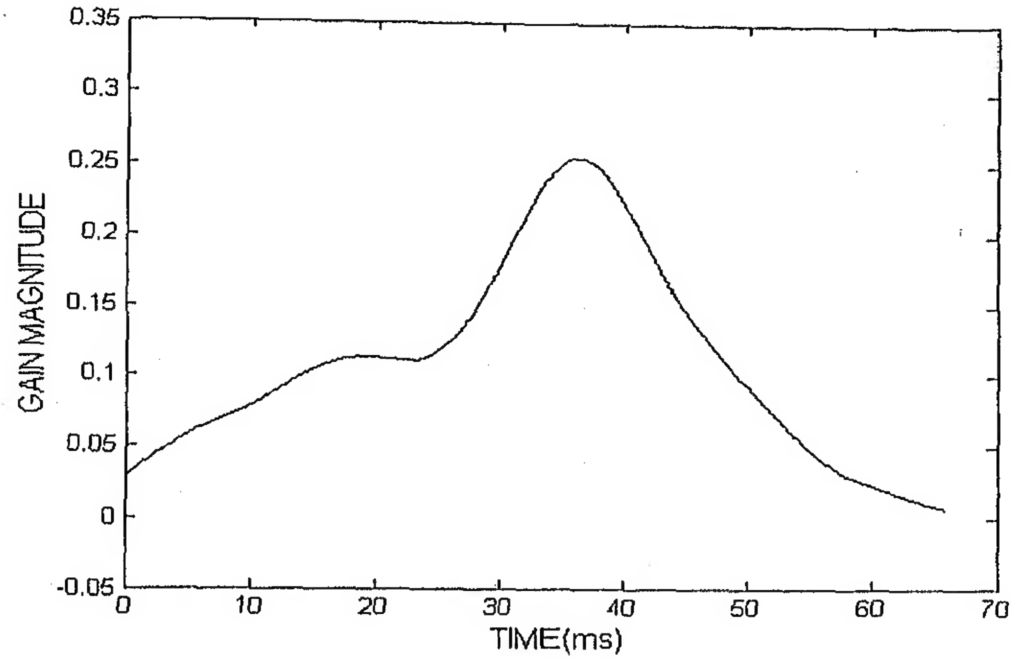


Fig .4.21: The original Gain function of phoneme sound 's'

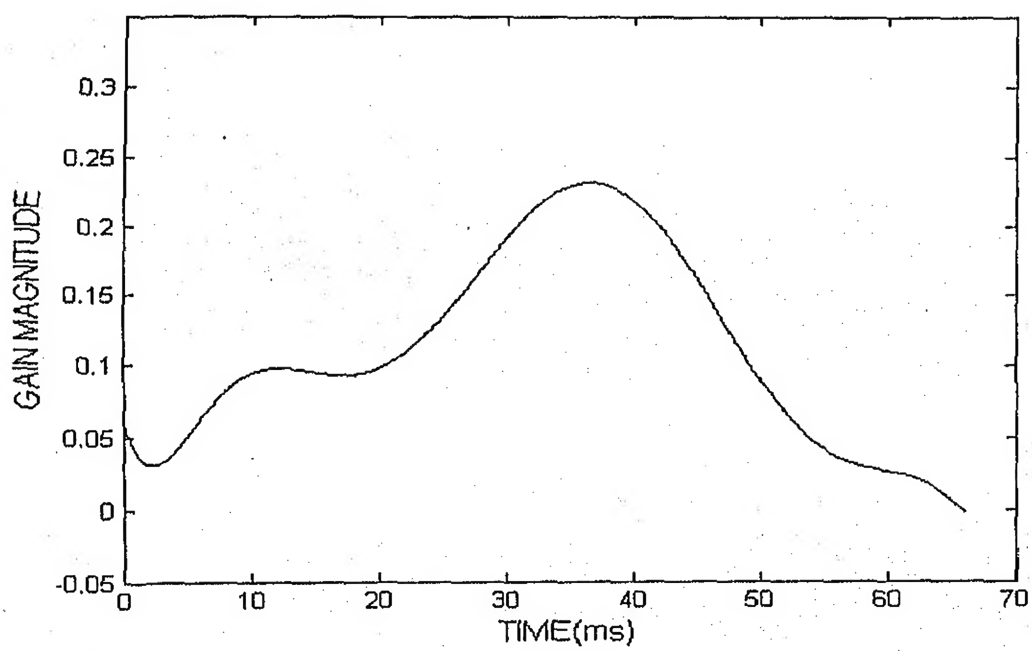
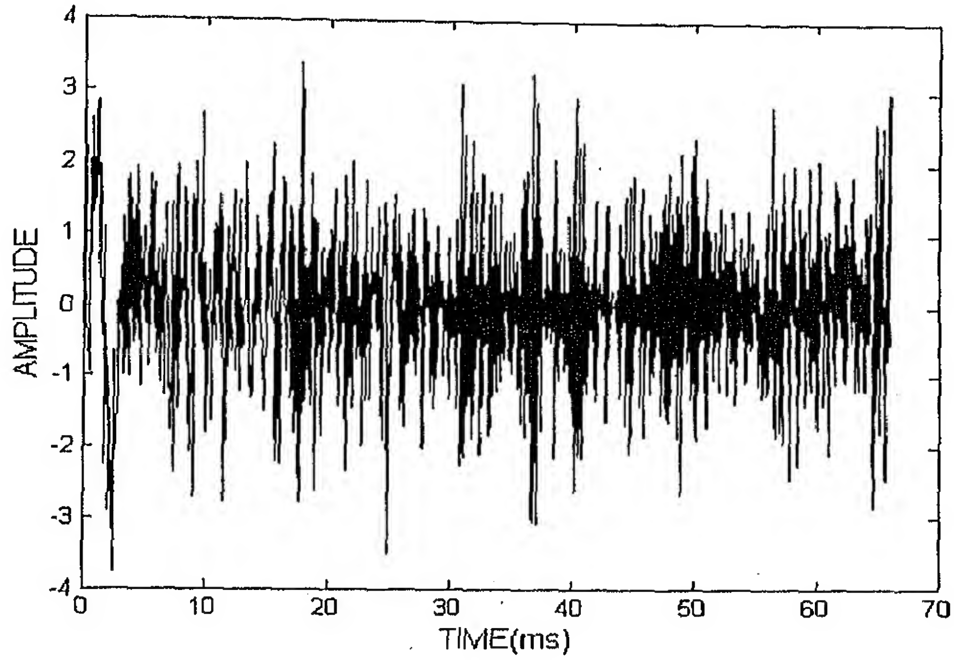
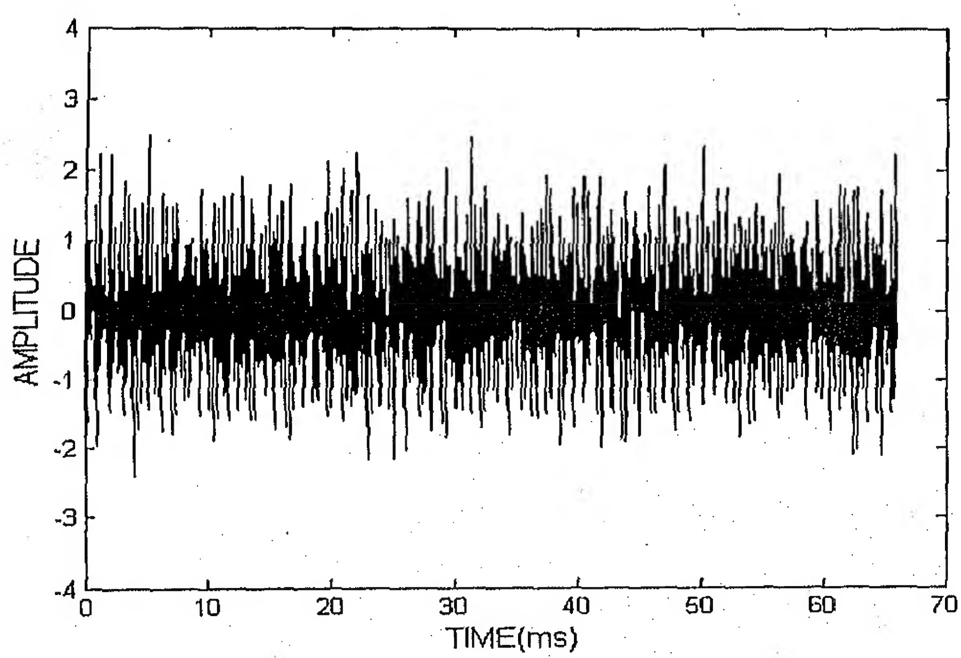


Fig.4.22: The regenerated Gain function of phoneme sound 's'



(a)



(b)

Fig.4.23: The phoneme sound 's' after division of gain function
(a) Original
(b) Regenerated

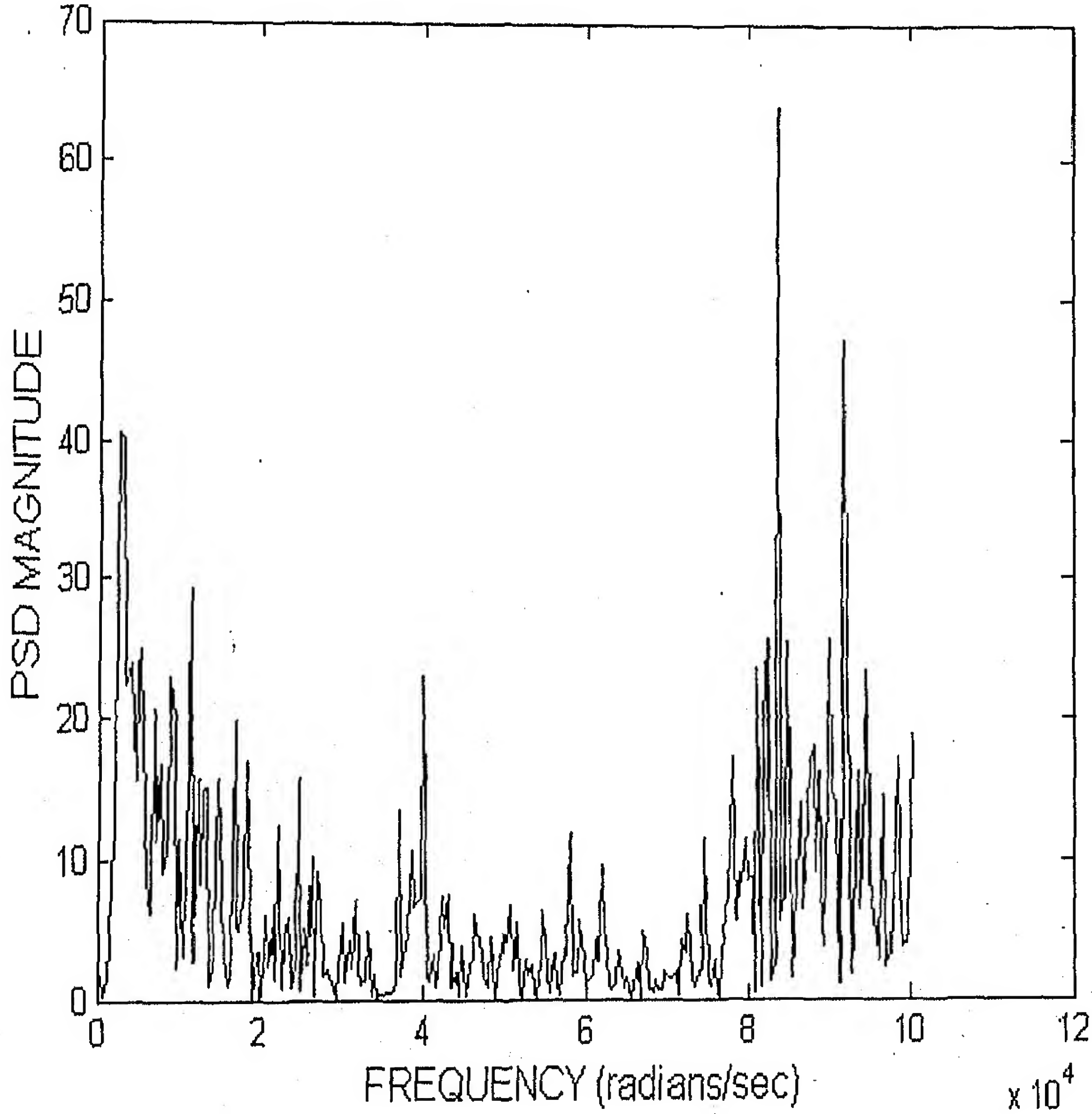


Fig. 4.24: The PSD plot of phoneme sound 's'

Table. 4.1: Separated words with respect to sample No.

SAMPLE #	WORD
2260 to 4600	don't
4600 to 8640	ask
8640 to 9520	me
9520 to 10736	to

Table.4.2: The phonetic Transcription

SAMPLE #	PHONEME
2260 to 2730	d
2730 to 4120	uh
4120 to 4600	n
4600 to 6864	ae
6864 to 7920	s
7920 to 8270	kcl
8270 to 8640	k
8640 to 8856	m
8856 to 9520	ix
9520 to 9960	dx
9960 to 10736	ix

Table 4.3: Estimated parameters for phoneme sound 'n'

#	ω	ν	μ	Amp (A)	Phase (ϕ)
1.	± 1408	± 2924	2.2989	0.3428	± 2.2449
2.	± 5714	± 1427	1.3315	0.2463	± 4.2662
3.	± 12804	± 1380	1.9458	0.1144	± 3.7867
4.	± 15649	± 6770	0.4805	0.1244	± 0.7324
5.	± 1433	-	-	0.7193	± 1.3445
6.	± 2876	-	-	1.5741	± 6.1514

Table. 4.4: Polynomial coefficients for gain function of phoneme sound 'n'

p_1	p_2	p_3	p_4	p_5
-1.1080×10^{-009}	1.0491×10^{-006}	-0.0003	0.0390	-2.4542

Table.-4.5: Estimated parameters for phoneme sound 'dx'

#	ω	ν	μ	Amp (A)	Phase (ϕ)
1.	± 1380	± 2641	0.1503	1.4753	± 1.3789
2.	± 5379	± 1342	1.4655	0.1388	± 0.0888
3.	± 11832	± 232	1.1430	0.1468	± 4.2269
4.	± 17264	± 410	1.1512	0.0874	± 5.4351
5.	± 2754	-	-	0.6329	± 3.9300

Table .4.6: Polynomial coefficients for gain function of phoneme sound 'dx'

p_1	p_2	p_3	p_4	p_5	p_6
-4.5462×10^{-012}	3.2055×10^{-009}	-1.1448×10^{-009}	-0.0003	0.0743	-3.2021

Table.4.7: Estimated parameters for phoneme sound 'ix'

#	ω	ν	μ	Amp (A)	Phase (ϕ)
1.	± 1449	± 1424	1.1573	0.1267	± 1.4196
2.	± 5835	± 1405	0.8056	0.1483	± 3.7086
3.	± 11487	± 1417	0.7755	1.1179	± 4.7595
4.	± 3007	-	-	0.4439	± 3.4297
5.	± 4358	-	-	0.5874	± 4.1344

Table.4.8: Polynomial coefficients for gain function of phoneme sound 'ix'

p_1	p_2	p_3	p_4	p_5	p_6	p_7
-5.008×10^{-015}	1.1707×10^{-011}	-9.6643×10^{-009}	4.1041×10^{-006}	-0.0009	0.0785	-1.1776

Table.4.9: Estimated parameters for phoneme sound 's'

#	ω	ν	β	Amp (A)	Phase (ϕ)
1.	± 24619	± 1330	0.70	0.5868	± 2.4613
2.	± 27162	± 470	8.40	0.4000	± 2.7800
3.	± 36918	± 2000	6.70	0.2610	± 5.9942
4.	± 39711	± 900	0.10	0.1572	± 0.5916
5.	± 47254	-	-	1.1028	± 3.2031

Table.4.10: Polynomial coefficients for gain function of phoneme sound 's'

p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9
2.92×10^{-021}	-1.39×10^{-017}	2.69×10^{-0}	-2.67×10^{-011}	1.45×10^{-008}	-4.20×10^{-006}	0.0006	-0.0292	-0.690

4.4 Quantization of Parameters

The quantization of various estimated parameters is explained in this section. For the purpose of explanation parameters of phoneme sound 'n' are taken.

By seeing the table (4.3) we find that total of 26 parameter values are required to synthesize the phoneme 'n'. Before quanitizing the frequencies ω_is and v_is are divided by 10000 and. After that mantissa and exponent parts are separately quantized. Mantissa part of each frequency can be quantized by $\approx$ 6 bits. There are total of 10 frequencies, hence, $\approx$ 60 bits are required to quantize the mantissa part of all the frequencies. 6 bits can quantize the exponent part. By this way total of 66 bits can quantize all the frequencies.

There are 4 modulation index parameters, 6 amplitude parameters and 6 phase parameters and to quantize each of these $\approx$ 6 bits are required which make a total sum of $16 \times 6 = 96$ bits to quantize all these parameters.

Polynomial coefficients for gain function of phoneme sound 'n' are shown in table (4.4). Here 1 bit is required for sign check, $\approx$ 6 bits are required to quantize mantissa part and $\approx$4 bits are suffice to

quantize the exponent part thus one coefficient can be quantized by ≈ 10 bits. There are 5 such coefficients, which can be quantized by ≈ 55 bits.

Thus we see that total number of bits required to quantize all the parameters of phoneme sound 'n' by simple uniform quanitization is $\approx$ 66+96+55, which is equal to 217 bits. And time interval of phoneme 'n' is 30ms. So coding rate turns out to be $\approx$ 7.2 Kbps.

In the similar way coding rate for other phoneme parameters can be approximately calculated. Here the method of quantization is very crude and is used only to check the potential of data compression .If other optimal methods of quantization like Vector quantization etc. are applied then more reduction in coding rate is anticipated.

Chapter 5

Conclusion

The phonemes having energy in low frequency region are well modeled by Complex AM, while phonemes having most of the energy in high frequency region are best modeled by complex FM model.

Also in this method of speech coding the parameters remain same for entire duration of time and approach is not frame based. Which reduces many complexities.

Due to change of parameter set at the phoneme boundary, there arises a problem of continuity while regenerating speech signal and further work in this direction is desirable.

Bibliography

[1] P.Sircar and M.S.Syali,"*Complex AM signal model for non-stationary signals*", Signal Processing, Vol 53, 1996, pp35-45.

[2] P.Sircar and S.Sharma, "*Complex FM signal model for non-stationary signals*", Signal processing, Vol 57, 1997, pp283-304

[3] P.Sircar and Major R.K.Saini, "*Parametric Modelling of Speech Signal,*" Thesis submitted to the department of Electrical Engineering , IIT Kanpur , Feb. 1998·

[4] S.M.Kay, "*Modern Spectral Estimation Theory and Applications*", Prentice Hall, Englewood Cliffs,NJ,1988.

[5] Gray, A.H. and Markel, J.D., " *Distance Measure for Speech Processing*", Proc. Of IEEE, Vol 81, no.9,pp1215-1247, Sep1993.

[6] V.C.Klema A.J.Laub, "*The singular value decomposition, its composition and some applications*", IEEE Trans.Autom.control, Vol 25, 1980, pp164-176.

[7] P.K.Sadasivan, D.Narayan Dutt, "*SVD based technique for noise reduction in electroencephalographic signals*", Signal Processing Vol 55,1996, pp179-189.

[8] W.H.Press, B.P.Flannery, S.A.Teukolsky and W.T.Vetterling, "*Numerical Recipes (FORTRAN Version)*",Cambridge University Press,1990.

[9] S.Haykin, "*Communication Systems*", Wiley,New York,1983.

[10] N.W.Mclachlan, "*Bessel Functions for Engineers*", Claredon Press, 1995.

[11] P.Sircar, S.mukhopadhyay, " *Modeling non-stationary signals by*

time-dependent AR process with time-varying gain", IETE Journal of Research, Vol 43,no.5, sept-oct 1997, pp 351-358

[12] Robert J. McAulay and Thomas F.Quatiery, " *speech analysis/synt--hesis based on a sinusoidal representation.* ", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol.ASSP-34, No.4, pp 744-754, Aug.1986

[13] A.Isaksson, A.Wennberg and L.H.Zetterberg, "*Computer analysis EEG signals with parametric models*", Proc.IEEE, Vol.69,April 1981, pp.451-461.